Query Match — DB 10

```
Query Match
Best Local Similarity   11.1%;   Score 784;   DB 10;   Length 1188;
                        29.9%;   Pred. No. 1.3e-44;
Matches 230;   Conservative 59;   Mismatches 359;   Indels 122;   Gaps 33;

Qy  119 PTPKPPPVVDEAGSGLDNGDFKVTTPDYSTTQOHNKVSTSPKITTAKPINPRPSL----PPN 174
Db  460 PTPHSPPAD-------DYVPPTPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQ 510

Qy  175 SDTSKETSLTVNKETTVETKETTTTNKQTSTPDGKEKTTISAKETQSIEKTSAKDLLAPTSKV 234
Db  511 APVG------------------SPPPPVKTTSPPAPIG---SPSSPPVSVSV 541

Qy  235 ----LAKPTPKAETTTKGPALTTPKEPTTPTPKEPASTTPKEPTTPTPKEPA 289
Db  542 SPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVASPPPPVKSP-PPPTLVASPPPPVKSPP 599

Qy  290 PTTTKSAPTTP--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTP-KEPAPTT 346
Db  600 PPAPVASPPPPVKSPPPPTPVASPPPPVASPPPPMKSPPPPTPVSSPPPPEKSPPPPP 659

Qy  347 PKKPAPTTPKEPAPTT-----PKE---PTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKP 397
Db  660 PAKSTPPPEEYPTPPTSVKSSPPPEKSLPPPTLLIPSPPQEKPTPPSPSKP-PSSPEKP 718

Qy  398 APTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTNKSAPTTTKEPAPTTTKSAPTTK 457
Db  719 SP--PKEPVSSPPQTP----KSSPPPAPVSSPPPTPVSSPPALAPVSSPPSVKSSP---- 768

Qy  458 EPSPTTTKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKPKPAPTAPKEPAP 517
Db  769 PPAPLSSPPPAPQVKSSPPPVQVSSP-PPAPKSSPPLAPVSSPPPQVEKTSPPPAPLSSPP 827

Qy  518 TTPKETAP-----TTP----KKLTPTTPEKLAPTTPEEPT 564
Db  828 LAPKSSPPHVVSSPPPPVVKSPPPPPLTPKPASPPAHVSSPPPPEVKPSTP--PA 885

Qy  565 PTTPEEPARPTTPKAAAPNTPKE-PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTT 623
Db  886 PTTVISP-PSEPKSSPPPPIVSLPPPIVKSSPPPAMVSSP-PMTPKSSPPPVVVSSPPT 943

Qy  624 LKE---PA----PPTPKK---PAPKELAPTTTKE--PTSTTSDKPAPTTPKGTAPTTP 669
Db  944 VKSSPPPAPVSSPPATPKSSPPAPVNLPPPEVKSSPPPPTPVSSPPPA---PKSSPPPAP 1000

Qy  670 -KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPT 728
Db 1001 MSSPPPPEVKSPPPPAPVSSPPPAPVSSPPPPVKSPPPPAP-----PPVKSPPPAPVSSPP 1057

Qy  729 -TPKETAPTTPKEPAPTTPKKPAPTTPEPTPPTTSEVSTPTTTKEPTTIHKSPDESTPEL 787
Db 1058 KSPPPPAPISSPPPEVKSPPPPAPVS--SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS 1115

Qy  788 SAEPTTPKALENSPKEPGVPTTKTPAAATKPEMTTTAKDKTTERDLRTTPEI 837
Db 1116 SPPPAP------VKPPSLP-PPAPVSSPPPVVTAPPKKEEQSLPPPAES 1158
```

Query Match — DB 11

```
Query Match
Best Local Similarity   11.0%;   Score 776.5;   DB 11;   Length 2187;
                        27.6%;   Pred. No. 7.7e-44;
Matches 290;   Conservative 119;   Mismatches 416;   Indels 227;   Gaps 48;

Qy   98 PDYESFCAEVKDNKKNRTKKKPTKPPPVVDEAGSGLDNGDFKVTTPDTSTTQOHNK-----152
Db  809 PKVDPIMSDVTPTSPKKTSATAVPK------------------DTSATLSLKSVPAV 847

Qy  153 VSTSPKITTAKPINPRPSLPPNSDTSKETSL--TVNKET-------TVETKETTTTNKQTS 204
Db  848 TSLSPKAPVAPSNEATIVPTEIPTSLKNALAAATPKETLATSISPKVTSPSPQKTPKSVS 907

Qy  205 TDGKEKTTISAKETQSIEKTSAKDLAPTSKVLAKPPPKAETTTKGPALTTPKEPTTPTPKE 264
Db  908 LKGAPAMTSKKAT---EIAASKDVSPSQ--FPKEVPLLQHV---PPTSPPKRSPSVSDTLSG 959

Qy  265 PASTTPKEPTPTTIKSAPTTPKE-PAPTTTKSAPTTPK------EPAPTTKEPAPT 314
Db  960 ALTSPPPKGPPPATLAEETTPTYPKKSPKPAASKKTPAPPSPEGVTAVPLEIPCSKRAPKTA 1019

Qy  315 TPKEPAPTTPKEPAPTT---TKSAPT------TPKEPAPTTPKK-P 350
Db 1020 APKESSAATSSSKRAPKTAVSKEIPSKGVTAVPLEISLPKETSKSATPGEKSASSPKRSP 1079

Qy  351 APTTPKEPAPTTPKEPTPTTPKEPAPTTTKEPAPTTPKEPAPTAPKKPAPTT--PKEPAP- 407
Db 1080 KTAGPKE---TPPGGVTAVPPEISLPKETPQNATPNESLAASSQKRSPKTSVPKETPPG 1136

Qy  408 --TTPKEPAPTTTKEPSPTTPK-----EPAPTTTKSAPTTTKEPAPTTT-KSAPTTPK- 457
Db 1137 GVTAMPLEIPSAPQKARPKTAVPKQIPTPEDAVTIILAGSPLSPKKASKTAAPKEAPATPSV 1196

Qy  458 ------EPSPTTTKEPAPTTPKGTAPTTPKE-PAPTTPKKPAPTT--PKE-----RAPTTP 496
Db 1197 GVIAVSGEISPSPKKPSPSKTAAPKENSATLPKKRSPKTAAPKETPATSSEGVTAVPSEISP 1256

Qy  497 KEPAP------------TTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPK 548
Db 1257 SPPTPASKGVPGVTLTPKGAPNALAE--SPASPKKVPKTAAPEETSTTPP--------PA 1309

Qy  549 PTTPEELAPTTPEEPTPTT------PEEPAPTTPKAAAPNTPKEPAPTTPKE-PAPTT 599
Db 1310 VAGPKEASATPPSKKIPKTAVPKETSAPSEGVTAVPLEIPPSPRKAPKTAAPKETPAPS- 1368

Qy  600 PKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTP----------KKPAPKE 639
Db 1369 -PEGATAPVQIPPSPRKGSKKAGSKE-TPTTPSPEGVTAAPLEIPISSKKTSKMASPKE 1426

Qy  640 LAPTTTKEPTS-------TTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTT 692
Db 1427 TLVTPSSKKLSQTVGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLTPSPK-DAPTT 1485

Qy  693 LKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKGTAPTTPKETAPTTPKKPAPT 752
Db 1486 LAE-SPSSPKK-APKTAAPPSER-VITVPPEKPA-TPQKASGTTIASKVPVPAETQEVAVS 1541
```

```
Qy    753  TPETP-------PPTTSEVSTPTTTK------EPTTIHKSPDESTPELSAEPTTPKALENSPK    801
                ||||       |:|               | :|    ||| |  |  | ::  |: |||
Db   1542  SRETPVTPAVPPVKNPSSHKKTSKTIELKEAPATLPPSPTIKSPKIPSSKKAPRT--SAPK   1599

Qy    802  E-PGVPTTKPPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKI    860
            |  :|:::|  ||  | |::| :   |  |   ||| |  || ||  | :|||| |: ||
Db   1600  EFPASPSIK-PVTT--SLAQTAPPSLQKAPSTTIPKENLAAPAV----LPVSSKSPAAPA   1652

Qy    861  TATTTQVTSTTTQDTTPFKIITLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAIN    920
            :| |  |: |  |  ||  :| | |||| | |||||| :|||| |:|||  ||| |||:
Db   1653  RASASLSPATKAAPQTPAPKEATTIPSCKKAAATETPIETSTPAPSLEGAPKETSE---TSV   1708

Qy    921  SKATTPKPQKPTKAPKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAML    980
            |||  | :| |  :|:|| | | || |:|||||:|| |||||| |||||:|| : | ||
Db   1709  SKVLMSSP-------PKKASSSKRASTLP------------ATTLPSLKEASVLS----   1744

Qy    981  QTTTRPNQTPNSKLVEVNPKSEDAGGAEGETP   1012
            :|:||  |  |:| |    ||:|| ||| |
Db   1745  -----PTATSSGKDSHISPVS-DACSTGTTTP   1770
```

RESULT   14

Q9VEL9

ID   Q9VEL9   PRELIMINARY;   PRT;   2112 AA.
AC   Q9VEL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG4090 PROTEIN.
GN   CG4090.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nixon K.A., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003716; AAF55402.1; -.
DR   FlyBase; FBgn0038492; CG4090.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF01607; Chitin_bind_2; 10.
DR   SMART; SM00494; ChtBD2; 11.
DR   PROSITE; PS00022; EGF_1; 1.
SQ   SEQUENCE   2112 AA; 219547 MW; B910185551A5D36 CRC64;

```
Query Match            10.6%;  Score 752;  DB 5;  Length 2112;
Best Local Similarity  26.6%;  Pred. No. 3.3e-42;
Matches  289;  Conservative 105;  Mismatches 356;  Indels 338;  Gaps 47;

Qy     78  FERGRCDCD------------AQCKKYDKC----------------CPDYESFCAEVKD-NKKNRT    115
               || | |           |  | |  |                   || ||: |:|  |||:|
Db    969  FKPAEKCESEETFLADNENCSKFYRCVDNGKCGFTKVSFTCPPNTLWDPEANSCNHPDQI   1028

Qy    116  KKKPTFKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPPNS    175
            :||   |                :  :|||:|||:::|||  |   |    ||  |   ||
Db   1029  QTKPLRCKKVVSQGGSSSN--------STSNSSSSSNNSGSSSNSGSSSS------SSSSNS   1076

Qy    176  DTSKETSLTVNKETTV---ETKETTTTNKQTS------TDGKEKTTTSAKETQSI------EK    222
             |:                  :||| ::||:|       ||||:|||| |||| :       ||
Db   1077  GSSSNTGSSSNSGASSSGGSSNQGSSSNSSGSNSSGGNQSTSSSTSSSSSSSNNNNQG   1136

Qy    223  TSAKDLAPTSKVLAKPTTPKAETTTKGP-------    249
            :|:    || |:|  |||| |:||||||
Db   1137  SSSSSSSSSSSSTSSSKPNPSETCKVNGQFIGDRSDCAKFYRCVDNDRGGFNMVFSCGPGT   1196

Qy    250  -------ALTTPKEPTTPTTPKEPASTTPKEPTTPTTIKSAPTTPKEPAP    290
                    |:|||||||||||| |||||||||| ||||:||||||||||||
Db   1197  VWDAQMQACNHAWAVKECGGIAPPTTPSTPTTSR-PTTASTSRPSDQTSTSRPTGP----P   1251

Qy    291  TTTTKSAPTTPKEPAPTTTTKEPAPTTPKEPAPTT----------------    323
            ||  ||||||||| || || ||  :|:  |  ||::
Db   1252  TTARPVTARPTTSSPTTASSSQTTSPVTQAPNTDGKCRSEGFMADPNNCSKFYRCVRNNK   1311

Qy    324  ---TKEP------APT------TTKSAPTTKREPAPTTPKKPAPTTPKEPAP    360
               ||||      |||       ||:||||:|:|||||||:||||||||||||
Db   1312  GGFTSIPFCQCGAGTVWDQLDQLQTCNHNFNNCSTGTESTTPKEPCC--EPATNGTTATSTSST   1369

Qy    361  TTPKEPTTPTTPKEPAPTTKEPAPTTAPKKPAPTTPKEPAPTTPKEPAPTTKE    420
            ||| |  :| |||||||||:| || |  ||||||||||||||||| |||||||
Db   1370  TTP--PPTTTDLPPTSTTGLP-PTTTTELPPT------TTTTDLPPTTTTRLPTTTTSL   1419

Qy    421  PSPTTPKEPAPTTTTKSAPTTTKEPAPTTTKSAPTTPK----EPSPTTTKEPAP------    469
             | ||||||| |||| ||| |: ||||||||:|||||     | |||||||||
Db   1420  PPTTTTGLPPTTTTGAQPTTTTLSSETETSTVTISPESTTQPPSTTTMKPLPAGTECTGE   1479

Qy    470  ----------------------TTPK---------------EPAP---TTPKKPAP    485
                                  |||                || |    :|||| ||
Db   1480  GYMADPEDCRKYYRCINAGASYRKYNFTCPKGTGWNEEVQTCDYVENIPRCSKLDPAEPIT   1539

Qy    486  TTP----KEPAPTTPKE-PAPTTVTKKPAPIAPKEPAPTTPKETAPTTPKKLITPTTPEKLA    540
            |||    |||||||||  |||:|  |||||  |||||||||||:|||||||  :|||||||
Db   1540  TTPSEESKDPGSTTPQSTTDEPTTVTKPI-TKPTE-EPSTEKPQKPTTQYPEKPTTTEE--   1595

Qy    541  PTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTT-    599
            | ||:||:||||| ||||:|| |||:||||:|||||||  ||  ||||||||||||||||
Db   1596  PEKPQKPT-TTEYPQKPTTTEEPTPEKPQKPT-TTEYPQKPTTTEEPTPEKPQKPTTTEY   1653

Qy    600  PKEPAPTTPKETA-----PTTPK--GTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTT    652
             ||| |:||::        |||   |  |||  :|||||| | ||| | |||||:|||:||
Db   1654  PQKPTTEEPTPTTSIPGYNPTTTSVPGYNPTTTPIPVETTTSTPGYK---PTTTGEPITTT   1710

Qy    653  SDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKK----PAP    706
             | ||| ||:|||||||||||||:||  |||||| |||||||  |||||||     |||
Db   1711  T-LPSTTTDAIQEPTTSKKPEPTTTTESPESSTPEGSVTTLQPEPQPNYNCSSEGFFPDP   1769
```

```
Qy   707 KELA------PTTKGPTSTSDKRAPTTPKETAPTTPKEPAPTTPETPPPT 760
         ::|:                  :| :|  |  |||  |  |  |     |
Db  1770 EDCSRYVRCVDAAKNGKYQVYAFKCGKGTVWDTSTETCNYADQVSGN-----CS 1818

Qy   761 TSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKPPAATKPEMTT 820
         :|:  :|| ||||  |     || :|        |      |          | | ||||
Db  1819 SGQTTPGTTTPEPGTTESTTSSGKP------ETTSKAPEN------TTTWA---PETTT 1862

Qy   821 TAKDKTTERDLRTTPEITTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKI 880
         : |    |      :|| |    |                      :           ||
Db  1863 T-----------SSPETTTT-----VASETTTTSGT------TTTATPETTTKPPKP--- 1898

Qy   881 TTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTS 940
         ::||:|  :| :| |||  |||||                    | ||  |    ||
Db  1899 --ETTTIAGEETSTSKSPTTTE-----------------SPAPSTNTSAP----- 1929

Qy   941 TKKPKTMP 948
          ||| |
Db  1930 --CPETGP 1935


RESULT 15
Q9XDH2

ID   Q9XDH2       PRELIMINARY;      PRT;      763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH MUCIN HOMOLOG.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang Y., Moreno C., Singh M.;
RT   "Cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003882; Pistil_extensin.
DR   PRINTS; PR01222; ATROPHIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSN.
SQ   SEQUENCE 763 AA; 39168EC45A5916F8 CRC64;

SEQUENCE FROM N.A.


Db   202 SPPRPPAP-----PMPATPMEFPPLPPVPPDDPISKETPPAPPAPPIPPAPVPIPPVPPLP 256
         : :|  ||     |  |             | :|  |||  |||  |  |  |    |||
Qy   526 TTPKKLTPTTPEKLAP------TTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTP-- 575

Db   257 PVPNKIPPAPP---APPVAVAAVLVARPCPPLPPLPNNHPAPPAAPVPGVPLAPLPNSHP 313
                |  || |   |                     ||    |    | ||| |
Qy   576 ---PKAAAPNTPKEPAP------------------------------------------- 589

Db   314 PAPPSAPVPGVPLAPLPISGRPVSVWKGSFTTLSTFCCRVCSGGEVLAGALNPSRPSRSPL 373
          ||| || | ||| ||| |  |  | |        ||              |   |  ||
Qy   590 --TTPKEPAPTTPKEPAPTTPKETA-----PTTPKGTA---PTTLKEPAPTTP--KKPAPKE 639

Db   374 TTTTPALPAPIPPLPLPPLPLPINTAVPPIPPLPPVTALAPPLPPLAPLPISPGVPRAP-- 431
         || |  |  | |||| ||| | |   |     | | |   ||| | ||| | | | ||
Qy   640 LAPTTTKEPTSTTSDKRAPTTPKGTAPTTPKEPA-----PTTPKEPAPTTPKGTAPTTL 693

Db   432 --PIPPGKPWTTPPLPAPAPPEPK-TVPVLPPGGPSCPPSEKPNPPAPPEPEPPERKSSPALPP 488
                  |        | | ||||   |  |   |      | |  |  |       |||
Qy   694 KEPAPTTP---KKPAPKELAPTTTKGP-TSTTSDKRAPTTPKET---APTTPKEPAPTTP 746

Db   489 APPAPSMPSAVRVPPSPPIPPAPPAAPRASMPALPAPPSPPATRLCPPLPPSPPAPNSP 548
         | || : | | | || | ||| |  |   |  || | ||| | |  || | ||| |||
Qy   747 KKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSP-----DESTPELSAEPTPKA--LENSP 800

Db   549 --PAPPAPPTPKLLS--ANPCPPVPPAPNRPPAPPAPPAPPPAPPELPAPPDPPPPPVANSP 604
             |  | | |       |  | | |    ||| | |        | | | ||   ||||
Qy   801 KEPGVPTTKPTPAATKPEMTTTAKDKTTERDLRTTPEITTTAAPKMTKETATTTEKTTESKI 860

Db   605 --PAPPAPPAPPSALPFVNPPA-----------PPTPAAPK------------SRPAL 637
                  |  |        |  |           |  | |||             |||
Qy   861 TATTTQVTSTTTQDTTPFKITTTLKTTTLAPKVTTTKKITITTTEIMNKPEETAKPKDRATN 920

Db   638 PAAPPAPPAPPVRATTP--------------------PPAPPAPPAPN 665
          |  |  |   |   ||                    ||  |  |||
Qy   921 SKATTPKPQKP------TKAPKKPTSTKKPKTMPRVRKPKTTPTPPRKMTSTM-----PELNP 971

Db   666 SMALPPAPPDPPTPLLATPPAPPAPPLPMSPPAPPLPPAAPDPPAPPLTINQPPSPPLAP 725
          : |  |  |   || | ||| | |  | | ||| | |  | || || | | ||| |||
Qy   972 TSRIAEAMLQTTTRPNQTPNSKLVEVENPKSEDAGGAEGET 1011

Db   726 VPGAPLAPLPINGRPVFARKNSLI------------GSSSGDT 756


Query Match
Best Local Similarity   28.2%;
                        10.1%;    Score 715;    DB 2;    Length 763;
Matches 248;    Conservative 39;    Mismatches 361;    Indels 232;    Gaps 39;
                        Pred. No. 3.5e-40;


Qy   238 PTPKAETTIKGPALTTPKEPTPTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTKSAP 297
         | | |    |   :||||| | ||||| ||||| |  | |||||||||||||   |
Db     3 PVP------APRALAPLPPAPPAPAEPKSKPPFFPAP--------PAPPCWMLVSAAP 46

Qy   298 TTPKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTTKKPAPTTPKE 357
          || |  |   | | |     ||  |      |||  |  |   |  |  | |
Db    47 PCP---PAP----PAPPKPKSKAPPPVPPAPARELAPPLP--PAP-----PEAPRE 90

Qy   358 PAPTTTKEPTPTTT----PKEPAPTTPKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK 411
          || |          |  | |  | |  |  | |   |  |  |   |  | |  | | |
Db    91 SRPALPCPCPPPVVIPDPPEPEAAPVPVRAPNSPPPPPFPPARKFVPAPPVP--PVPNSPP 148

Qy   412 EP--APTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKEPSPTTTKEPAP 469
               | | | | || | | | |   |   | | | |   | | | | |  |
Db   149 FPPFPPAALNPPAP----PAPPLANSPPLPPAPPPPAGT--PPAAPWPPVPPAARKSKPA 201

Qy   470 TTPKEPAPTTPKPAPTTPKEPAPTTPKEPAP-----TTTKKPAPTAPKEPAPTTPKETAP 525
```

Search completed: April 26, 2002, 16:30:35
Job time: 650 sec

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: April 26, 2002, 16:29:54 ; Search time 210.7 Seconds
              (without alignments)
              944.835 Million cell updates/sec

Title: AA7
Perfect score: 7323
Sequence: 1 MAWKLPIYLLLLSVFVIQ.........ARAITTRSGQTLSKVWYNCP 1361

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_17:*
              1: sp_archea:*
              2: sp_bacteria:*
              3: sp_fungi:*
              4: sp_human:*
              5: sp_invertebrate:*
              6: sp_mammal:*
              7: sp_mhc:*
              8: sp_organelle:*
              9: sp_phage:*
             10: sp_plant:*
             11: sp_rodent:*
             12: sp_virus:*
             13: sp_vertebrate:*
             14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7291.5	99.6	1404	4	Q92954	homo sapien
2	7276.5	99.4	1404	4	Q9VPG1	homo sapien
3	3826.5	52.3	1054	11	Q9BX49	homo sapien
4	1713.5	23.4	401	6	Q9JM99	mus musculu
5	982	13.4	1079	5	O77765	bos taurus
6	955	13.0	1049	5	Q9N4S7	caenorhabdi
7	868.5	11.9	1795	5	Q917S1	drosophila
8	867	11.8	1489	5	O76894	drosophila
9	839.5	11.5	1315	10	O96449	phytophthor
10	835	11.4	1274	5	Q9SPM0	zea mays (m
11	830	11.3	1480	10	O20007	caenorhabdi
12	828.5	11.3	2187	11	Q9LIE8	arabidopsis
13	800	10.9	1188	10	P70670	mus musculu
14	757.5	10.3	2112	8	O41805	zea mays (m
15	715	9.8	763	2	Q9XDH2	drosophila
16	707.5	9.7	555	10	Q9FPQ6	chlamydomon
17	706	9.6	677	5	Q9N435	caenorhabdi
18	704.5	9.6	1151	13	O57580	gallus gall
19	688	9.4	4880	11	Q9JLT1	rattus norv
20	688	9.4	5085	11	Q9JKS6	rattus norv
21	687.5	9.4	2284	5	Q9VPG1	drosophila
22	674	9.2	1229	5	O94185	drosophila
23	670	9.1	7962	4	Q10465	homo sapien
24	668	9.1	3507	5	Q23587	caenorhabdi
25	663	9.1	3570	4	Q99552	homo sapien
26	659.5	9.0	4833	11	Q90YX6	mus musculu
27	659.5	9.0	5038	11	Q90YX7	mus musculu
28	657.5	9.0	2089	4	O14676	homo sapien
29	654	8.9	2768	5	Q9VC00	drosophila
30	653.5	8.9	1514	5	Q9GUW7	leishmania
31	646	8.8	990	13	Q91803	xenopus lae
32	639	8.7	1612	5	Q9VYQ2	drosophila
33	633	8.6	489	10	Q41707	vigna ungui
34	633	8.6	6632	5	O17362	caenorhabdi
35	632	8.6	761	10	Q9ZQI0	arabidopsis
36	632	8.6	971	5	Q9XVS4	caenorhabdi
37	626	8.5	6642	5	O01761	caenorhabdi
38	622.5	8.5	839	2	Q9RX57	deinococcus
39	620.5	8.5	2344	5	O9N3Y8	caenorhabdi
40	607.5	8.3	801	5	O23635	caenorhabdi
41	607.5	8.3	924	12	Q99307	epstein-bar
42	605	8.3	379	5	O27929	drosophila
43	600.5	8.2	1893	5	Q9NKC9	drosophila
44	598.5	8.2	409	10	Q9SBM1	volvox cart
45	577	7.9	956	10	Q9LJ64	arabidopsis

ALIGNMENTS

RESULT   1
Q92954

```
ID   Q92954      PRELIMINARY;      PRT;   1404 AA.
AC   Q92954;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEGAKARYOCYTE STIMULATING FACTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Blood 78:279-279(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA   FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA   Jacobs K., Turner K.;
RA   (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL   Mosher D.F. (eds.);
RL   Biology of vitronectins and their receptors., pp.45-52,
RL   Elsevier Science Publishers B.V. (1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
     EMBL; U70136; AAB09089.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   InterPro; IPR002400; GF_cysknot.
DR   Pfam; PF00045; hemopexin; 2.
DR   Pfam; PF00045; GF_cysknot.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR00438; GFCYSKNOT.
```

```
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
SQ   SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match            99.6%;  Score 7291.5;  DB 4;  Length 1404;
Best Local Similarity  96.9%;  Pred. No. 0;  Mismatches 0;  Indels 43;  Gaps 1;
Matches 1361;  Conservative 0;  Mismatches 0;

Qy    1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60

Qy   61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 120

Qy  121 PPSGASQTIKSTTKRSPKPPNKKKTKVIESEEITE------------------- 156
         |||||||||||||||||||||||||||||||||||
Db  121 PPSGASQTIKSTTKRSPKPPNKKKTKVIESEEITEEHSVSENQESSSSSSSSSSTIW 180

Qy  157 --------------VKDNKKNRTKKKPFTPKPPVVDEAGSGLDNGDFKVTTPDTST 197
                       ||||||||||||||||||||||||||||||||||||||||||
Db  181 KIKSSKNSAANRELQKKLKYVKDNKKNRTKKKPFTPKPPVVDEAGSGLDNGDFKVTTPDTST 240

Qy  198 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG 257
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG 300

Qy  258 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPPTTPKEPAS 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTTKGPALTTPKEPPTTPKEPAS 360

Qy  318 TTPKEPTTPTTIKSAPTTPKEPAPTTPKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTKEP 377
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTPKEPTTPTTIKSAPTTPKEPAPTTPKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTKEP 420

Qy  378 APTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 437
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 APTTPKSAPTTPKEPAPTTPKEPAPTTPKEPPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 480

Qy  438 EPAPTTPAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKSAPTTTIKEPAPT 497
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 EPAPTTPAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPT 540

Qy  498 TTKSAPTTPKEPSPTTTKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTKKP 557
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTTKSAPTTTKEPAPT 600

Qy  558 APTAPKEPAPTTPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEELAPTTPEETPTT 617
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEELAPTTPEETPTT 660

Qy  618 PEEPAPTTPKAAANPKEPAPTTPKETAPTTPKEPAPTTPKEKLAPTTPEELAPTTTKGTAPTTLKEP 677
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 PEEPAPTTPKAAANPKEPAPTTPKEPAPTTPKEKPAPTTPEELAPTTTKGTAPTTLKEP 720

Qy  678 APTTPKKPAPKELAPTTPKEPTSTTSDKPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKG 737
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 APTTPKKPAPKELAPTTPKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKG 780

Qy  738 TAPTTLKEPAPTTPKKPAPKELAPTTPKKGPTSTTSDKPAPTTPKEPAPTTPK 797
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 TAPTTLKEPAPTTPKKPAPKELAPTTPKGPTSTTSDKPAPTTPKEPAPTTPK 840

Qy  798 KPAPTTPEEPPPTTPSEVSPTTTHKPTTTHKSPDESTPELSAEPTPKALENSPKEPGVPT 857
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 KPAPTTPEEPPPTTSEVSPTTTKEPTTTHKSPDESTPELSAEPPKALENSPKEPGVPT 900

Qy  858 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 917
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 960

Qy  918 TSTTTQDTTPFKITTLKTTTLAPKVTTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 977
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 TSTTTQDTTPFKITTLKTTTLAPKVTTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 1020

Qy  978 PQKPTKAPKKPTSTKKPKTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1037
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 PQKPTKAPKKPTSTKKPKTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080

Qy 1038 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1097
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1140

Qy 1098 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSSPARRITEVWGIPSPIDTVFT 1157
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSSPARRITEVWGIPSPIDTVFT 1200

Qy 1158 RCNCEGKFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLLTGQIVAALSTAKYKNWPESVY 1217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 RCNCEGKFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLLTGQIVAALSTAKYKNWPESVY 1260

Qy 1218 FFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1277
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 FFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1320

Qy 1278 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV 1337
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV 1380

Qy 1338 PSRTARAITTTRSGQTLSKVWYNCP 1361
         |||||||||||||||||||||||||
Db 1381 PSRTARAITTTRSGQTLSKVWYNCP 1404
```

RESULT   2
Q9BX49

```
ID   Q9BX49;                    PRELIMINARY;      PRT;     1404 AA.
AC   Q9BX49;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).
GN   BG174L6.2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133553; CAC36090.1; -.
SQ   SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;

Query Match            99.4%;  Score 7276.5;  DB 4;  Length 1404;
Best Local Similarity  96.7%;  Pred. No. 0;  Mismatches 3;  Indels 43;  Gaps 1;
Matches 1358;  Conservative 0;  Mismatches 0;

Qy    1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60

Qy   61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 120

Qy  121 PPSGASQTIKSTTKRSPKPPNKKKTKVIESEEITE------------------- 156
         |||||||||||||||||||||||||||||||||||
Db  121 PPSGASQTIKSTTKRSPKPPNKKKTKVIESEEITEEHSVSENQESSSSSSSSSSTIR 180
```

```
Qy    157 -----------------------------VKDKKKNRIKKKPPTKPKPKPVVDEAGSGLDNGDFKYTTPDTST  197
                                       |||||||||||||||||  ||||||||||||||||||||||||
Db    181 KIKSSKNSAANRELQKKLKVKDNKKNRIKKKPPTKPPPVVDEAGSGLDNGDFKYTTPDTST  240

Qy    198 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLIVNKETTVETKETTTNKQTSTDG  257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLIVNKETTVETKETTTNKQTSTDG  300

Qy    258 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS  317
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS  360

Qy    318 TTPKEPTPTTIKSAPTTPKEPAPTTTKEPAPTTTKEP...PTTPKEPAPTTTKEP  377
          |||||||||||||||||||||||||||||||||||
Db    361 TTPKEPTPTTIKSAPTTPKEPAPTTTKEPAPTTTKEP...PTTPKEPAPTTTKEP  420

Qy    378 APTTTKSAPTTPKEPAPTTTKEPAPTTTKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP  437
          ||||||||||||||||||||||||||||||
Db    421 APTTTKSAPTTPKEPAPTTTKKEPAPTTTKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP  480

Qy    438 EPAPTTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTTTPKEPAPTTTKSAPTTTKEPAPT  497
          |||||||||||
Db    481 EPAPTAPKKPAPTTPKKEPAPTTPKEPAPTTTKEPSPTTTPKEPAPTTTKSAPTTKEPAPT  540

Qy    498 TTKSAPTTPKEPSPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKKP  557
          |||||||||||||||||
Db    541 TTKSAPTTTKEPSPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKKP  600

Qy    558 APTAPKEPAPTTPKETAPTTPKKLTPTTTPKKLTPTTPEKLAPTTTPEKPAPTTTKEPTTT  617
          |||
Db    601 APTTPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTTPEKLAPTTTPEKPAPTTTKKP  660

Qy    618 PEEPAPTTPKAAAPNTPKEPAPTTPKKLTPTTTPEKLAPTTTPEELAPTTPEEPTPTT  677
          |||
Db    661 PEEPAPTTPKAAAPNTPKEPAPTTPKKLTPTTTPEKLAPTTTPEELAPTTPEEPTPTT  720

Qy    678 APTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTKEPAPTTTPKG  737
          |||||||||||||
Db    721 APTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTKEPAPTTTPKG  780

Qy    738 TAPTTLKEPAPTTPKKPAPKELAPTTKIKGPTSTTSTRDKPAPTTPKETAPTTPKEPAPTTPK  797
          |||||||||||||||||
Db    781 TAPTTLKEPAPTTPKKPAPKELAPTTTKIKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPK  840

Qy    798 KPAPTPTPETPPPPTTSEVSTPTTPTTKEPTTIIHKSPDESTPELSAEPTPKALENSPKEPGVPT  857
          ||||||||||
Db    841 KPAPTTPETPPPTTSEVSTPTTPTTHKSPDESTPELSAEPTPKALENSPKEPGVPT  900

Qy    858 TKTPAAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV  917
          |||||||||
Db    901 TKTPAAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV  960

Qy    918 TSTTTQDTTPFKIITTLKTTITKKTITITTEIMNKPEETAKPKDRATNSKATTPK  977
          |||||||
Db    961 TSTTTQDTTPFKIITTLKTTTTLAPKVTTTKKTITITTEIMNKPEETAKPKDRATNSKATTPK  1020

Qy    978 PQKPTKAPKKPTSTKKPKTIMPVRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN  1037
          |||||||||||
Db   1021 PQKPFKAPKKPTSTKKPKTMPVRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN  1080

Qy   1038 QTPNSKLVEVNPKSEDAGGAEGEPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS  1097
          ||||||||
Db   1081 QTPNSKLVEVNPKSEDAGGAEGEPHMLLRPHVFMPEVITPDMDYLPRVPNQGIIINPMLS  1140

Qy   1098 DETNICNGKRPVDGLITTLRNGTLVAFRGHYFWMLSPFSPPSSPARRITEVWGIPSPIDTVFT  1157
          |||||||
Db   1141 DETNICNGKRPVDGLITTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFT  1200

Qy   1158 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLIGQIVAALSTAKYKNWPESVY  1217
          ||||||||||||
Db   1201 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLIGQIVAALSTAKYKNWPESVY  1260

Qy   1218 FFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHIRIQY  1277
          |||||||||||
Db   1261 FFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHIRIQY  1277
```

```
Db   1261 FFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHIRIQY  1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1278 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV  1337
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV  1380

Qy   1338 PSRTARAITTRSGQTLSKVWYNCP  1361
          ||||||||||||||||||||||||
Db   1381 PSRTARAITTRSGQTLSKVWYNCP  1404
```

RESULT 3
Q9JM99

```
ID   Q9JM99        PRELIMINARY;      PRT;   1054 AA.
AC   Q9JM99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE   PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN   PRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ikegawa S., Nakamura Y.;
RT   "a novel mouse gene highly homologous to a human gene encoding
RT   megakaryocyte stimulating factor precursor and cartilage superficial
RT   zone protein.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034730; BAA92310.1; -.
DR   MGD; MGI:1891344; Prg4.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001212; Somatomedin_B.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00024; SOMATOMEDIN_B; 2.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ   SEQUENCE   1054 AA;  115991 MW;  4FC64BFA42283235 CRC64;

Query Match           52.3%;  Score 3826.5;  DB 11;  Length 1054;
Best Local Similarity 54.9%;
Matches  776;  Conservative   59;  Mismatches  167;  Indels   ;  Gaps  24;
                                   Pred. No. 1.1e-245;
```

```
Qy      1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF   60
          | |||| | ||||  |||||||||||||||||||||||||||||||||||||||||||||
Db      1 MGWKILPVCLSLLLPVVLIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF   60

Qy     61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP  120
          ||||  ||||||||||| ||||||||| ||  | ||| ||||  |||| || ||| : ||
Db     61 KRVCSPELSCKGRCFESFARGRECDCDSQCKOYGKCCADYDSFCEEVHNSTS-PSSKTAP  119

Qy    121 PPSGASQTIKSTTKRSPKPNNKKTTKKVIESEEITE-----------------------  156
          | ||| | |||||||||| | :|  : | |||||:||
Db    120 TPAGASDTIKSTTKRSPKSPI-TRTIKVESEELTEEHSDSENQESSSSSSSSTIRKI    178

Qy    157 ------------VKDNKKNRTKKKPPTKPKPVVDEAGSGLDNGDFKVT--TPDTSTT    198
                       ||||||||| ||| ||||||:||||||||||| |:  || : ||
Db    179 KSSKNSANRELQKNPNVKDNKNKNTPKKKPNPEPPAVDEAGSGLDNGEFKLTPPPPDPPTT  238

Qy    199 QHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLIVNKETTVETKETTTNKQTSTDGK   258
          :||||:||||| ||| | | |||  ||:||| ||| :|||||||||||| |||:: : |
Db    239 PHSKVATSPKITAAKPVTPKPSLAPNSETSKEASLASNKETTVETKETTATNKQSSA-SK  297

Qy    259 EKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAST  318
```

```
Db  298  KKTTSVKEPRSAEKTSDKDV-------------------EPTST  322
         || || |||| :| |||| |:
Qy  319  TPKEPTTPTTIKSAPTTPKEPAPTTTKEPAPTTTKEPA  378

Db  323  TPK-----------NSAPTTKEPV-TTTKESKFLP-  360
         |||           |:|| |:|  |||| :|
Qy  379  PTTTKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKE  438

Db  361  PTTTKPEPTTPKREPEPTTPKEPEPTTPKEPEPTTPKE  420
         |||||:||||:|:|||||||||||||||||||||||
Qy  439  PAPTTAPKKPAPTTPKEPAPTTPKEPSPTTPKEPAPTTTKSAPTTTKEPAPTT  498

Db  421  PGPTTPKEPEPTTPKEPEPTTPKEPESTTPKEPEPTTPKEP-  461
         | |||||||||||||||||||||||:||||||||||||
Qy  499  TKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTTKEPAPTTTKKPA  558

Db  462  --EPTTPKEPEPTTLKEPEPTTPKEPEPTTPKEPEPTT-----  512
           |||||||||||| ||||||||||||||||||||||
Qy  559  PTAPKEPAPTTPKETAPTTPKKLIPTTPKEPEPTTPEELAPTTPEEPTPTTP  618

Db  513  -----------------TPKEPEPTTPKEPEPTTPKEPEPTTP----  542
                           ||||||||||||||||||||||||
Qy  619  EEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPA  678

Db  513  -----------------------KEPE  542
                                 ||||
Qy  679  PTTPKKPAPKELAPTTPKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKGT  738

Db  543  PTTPKKP--------------------------EPTTPKEPVPTTPKEPEPTTP--  570
         ||||||                          ||||||| |||||||||||
Qy  739  APTTLKEPAPTTPKKPAPKELAPTTPKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKK  798

Db  571  ----------KEPEPTTPKEPEP-  612
                   ||||||||||||
Qy  799  PAPTTPEPTPPPTTSEVSIPPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTT  858

Db  613  PEPTT---  617
         | |||
Qy  859  KTPAAATKPEMTTPAKDKTTPKDLPKTTPEPTTAAPKMTKETATTTEKTTESKITAFTTQVT  918

Db  618  --------TSPKT-  622
                 ||| |
Qy  919  STTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKAT----  974

Db  623  -----------TTLKATTLAPKVTAPAE--  665
                    |  :  |: ||||||||
Qy  975  ----TPKP-QKPIKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAM  1029

Db  666  KPTKAPKPKTPKKPTKAPKKPTSTKKPKT-PKTRKPKTTPAPLKTTSATPELNTTP--LEVM  722
         | ||||||||||||||||||||||| |||||||||||||||: ||| | | |
Qy  1030 LQTTTRPNQTPNSKLVEVNPKSEDAGGAEGEGTPHMLLRPHVFMPEVTPDMDYLPRVPNQG  1089

Db  723  LPTTTIPKQTPNPETAEVNPHDHEDADGGEGEKP-LIPGPPVLFPTAIPGTDLLAGRLNNG  781
         |||||| ||| |: ||  | |||| ||||  | | ||| |||||: || | || | |
Qy  1090 IIINPMLSDETNICNGKRPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPSPARITEVWGIP  1149

Db  782  ININPMPSDETNLCNGKPVDGLTTLRNGTLVAFRGHYFWMLNPFRPSPSPPRRITEVWGIP  841
         |:|||| ||||| |||| ||||||||||||||||||||||| ||  || || ||||||||
Qy  1150 SPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKY  1209

Db  842  SPIDTVFTRCNCEGKTFFFKDSQYWRFTNDVVDPGYPKQIVKGFGGLTGKIVAALSTAKY  901
         |||||||||||||||||||||||||||||:|:|||||:|:|||||||||:|||||||||
Qy  1210 KNWPESYVFFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQYRRRRFE  1269

Db  902  KDRPESVYFFKRGGNIQQYTYKQEPMKKCTGRRPAINYSVYGEAAQVRRRRFERAVGFPQ  961
         |: |||:||||||| |||| |||||:||| ||||:||:||||  ||||||||
Qy  1270 THTIRIQYS-PARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFS  1328

Db  962  THTFRIHYSVPMRVSVQDKGFLHNEVKVSTLWRGLPNVVTSAITLPNIRKPDGYDYYAFS  1021
         ||| || ||  :|  |||| ||||||||| ||||||||||||:||||||||||||||||
Qy  1329 KDQYYNIDVPSRTARAITTRSGQTLSKVWYNCP  1361
         ||| |||||||||||||||||||||||:|||||
Db  1022 KDQYNIDVPSRTARAITTRSGQTLSKIWYNCP  1054
```

RESULT 4

ID   O77765   PRELIMINARY;   PRT;   401 AA.
AC   O77765;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ARTICULAR CARTILAGE;
RA   Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA   Aydelotte M.B.;
RT   "Immunodetection and Partial cDNA Sequence of the Proteoglycan,
RT   Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT   Joints.";
RL   J. Orthop. Res. : 0:0-0(1998).
DR   EMBL; AF056218; AAD13404.1; -.
DR   Pfam; PF00045; hemopexin; 2.
DR   InterPro; IPR000585; Hemopexin.
DR   PROSITE; PS00024; HEMOPEXIN; 2.
DR   SMART; SM00120; HX; 2.
FT   NON_TER   1   1
FT   NON_TER   401   401
SQ   SEQUENCE   401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;

Query Match   23.4%; Score 1713.5; DB 6; Length 401;
Best Local Similarity   80.2%;
Matches 320; Conservative   24; Mismatches   52; Indels   3; Gaps   3;

```
Qy   964  PKDRATNSKATTPKPQKPTRAPKKPTSTKKPKTMPRVRKPKTTPTPRK-MTSTMPELNPT  1022
          |||||||| : ||||||||| ||||||||||||:|||||||||||||||  : | |||||
Db     5  PKGRATNSQVTTPKPQKPTKAPKKPTSTKKPRT-PRVRKPKTTPTPPKITTISAMPEPTPT    63

Qy  1023  SRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGEGTPHMLLRPHVFMPEVTPDMDYL  1082
          : ||||||||||| | |||| : :| || ||: || || ||| :|||| : || | | :
Db    64  S-LPEAMLQTTTRPPTPTPNSEIIDVNSENEDGDAAEGEKPHMIFRPPVLFPIVIPGTEII   122

Qy  1083  PRVPNQGIIINPMLSDETNICNGKRPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPSPARI  1142
          :|:|  |:|||| ||||| |||| |||||||||||||||||||||||||: ||: || |||
Db   123  VRGPSGGFGINPMFSDETNLCNGRPVDGLTTLRNGTLVAFRGHYFWMLPFTPPPPPRRI    182

Qy  1143  TEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVA  1202
          |||||||||||||||||||||||||||||||||||||||||||||: | :|||||:| :|
Db   183  TEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKLISKGFGGLNGKIVA   242

Qy  1203  ALSTAKYKNWPESYVFFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQYRRRRFE  1262
          ||| |:|| |||:||||||||:|||| |||||.||| ||||:||: || | || ||||||
Db   243  ALSIAQYKSRPESVYFFKRGGSVQQYTYKQEPTQKCTGRRPAINYSVYGETAQVRRRRFE   302

Qy  1263  RAIGPSQTHTIRIQYSPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGY  1322
          |||||||.|||||:|:| :| ||||| ||||||||:|||||||||||||||||||||||
Db   303  RAIGPSQVHTIRIHYTPVRVPYQDKGFLHNEVKVSTLWRGLPNVVTSAISLPNIRKPDGY   362

Qy  1323  DYYAPFSKDQYYNIDVPSRTARAITTRSGQTLSKVWYNCP  1361
          |||| :||||||||||||||||||||||||||| |||||
Db   363  DYYALSKDQYYNIDVPSRTARAITTRSGQTLSNTWYNCP   401
```

RESULT 5

ID   Q9N4S7   PRELIMINARY;   PRT; 1079 AA.
AC   Q9N4S7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

```
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y51B11A.1 PROTEIN.
GN   Y51B11A.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RA   Cotton M.;
RT   "The sequence of C. elegans cosmid Y51B11A.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   1079 AA;  110532 MW;  8DBDE3824CF80CA1 CRC64;


Query Match           13.4%;  Score 982;  DB 5;  Length 1079;
Best Local Similarity 28.7%;  Pred. No. 4.2e-57;
Matches  346;  Conservative  151;  Mismatches  501;  Indels  208;  Gaps  49;

Qy     38 YSR-DATCNCDYNCQHYMECCCPDFKRVCTAELSCGKGRCFESFERGRECDCDAQCKKYDKC 96
             ||| :
Db      2 WSRTDSPWACKFNRKHF------DQERAAIGQKELE----------IADCRANQTPHIML 45

Qy     97 CPDYESFCAEVHNPTSEPPSSKKAPPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITE 156
             || : ||
Db     46 PSTLTSVDME-----TPSTLVLSSTPTSSSTPIKETTTTAP------------ETTS 85

Qy    157 VKDNKNKNRTFKKPPKPPVVDEAGSGLDNGDFKVTTPDTSTTIQHNKVSTSP-KITTAKPI 215
             ||
Db     86 TEEPPSSSTTPVQTT-----------TTTAPETTSTEAPSSSSTTPVQTTTTTAP 127

Qy    216 NPRPSLPPNSDTS--KETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKT 273
             ||| : |
Db    128 ETTSTEPPSSSTTPVQTTTTTTAPETTSTEAPSSSTTPVQTT----------TTTAPETTSTEPP 181

Qy    274 SAKDLAPTSKVLAKPTPKAETTIKGP--ALTTKPEPTPTTPKEPASTTPKEPTTTIKSA 331
              : |
Db    182 SSS----TSPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTT 237

Qy    332 PTTPKEPAPTTKSAPTTPKEPAPTTKEPAPTTKEPAPTTTKSAPTTPKE 391
             ||
Db    238 TTTAPETTSTEPSSSTTPVQTTTTTAPE---TTSTEP-PSSSTTPVQTTTTTAPETTST 293

Qy    392 PAPTTPKPAPTTPKEPAPTTPKEPTTPKEPTTPTTPKEPAPTAPKKPAPT 450
             |||
Db    294 EPPSSSTTPVQTTTTTAPETTSTEP-PSSSTTPVQTTTTTAPETTSTEPPSSSST 346

Qy    451 TPKEPAPTTPKEPAPTTPKEPAPTTKEPAPTTTKSAPTTPKE 508
             ||
Db    347 TPVQTTTTT----APETRTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTA 402

Qy    509 PSPTTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKKPAPTAPKEP--A 566
             |:|:
Db    403 PETTSTEPPSSST--TPVQTTTTTAPETTSTEP-PSSSTTPVQTTTTTAPETTSTEPPSS 459

Qy    567 PTTPKETAPTTPKKLTPTTPEKPAPTTPEKPAPTTPEELAPTTPEEP--TPTTPEEPAPT 624

Db    460 STTPVQTTTTTAPETTSTEP------PSSSTTPVQTTTTTTAPETTSTEPPSS 515
Qy    625 TPKAAAAPNTPKEPAP---TTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTT 681
Db    516 T----APETTSTESPSSSTTPVQTTTTTAPETTSTEPPSSS-TTPVQTTTTTAPETTSTE 570
Qy    682 PKKPAPKELAPTTTKEPTTSTSDKP-APTTPKGTAPTTKEPAPTTPKEP--APTTPKGT 738
Db    571 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPE---TTSTEPPSSSTTPVQT 627
Qy    739 APTTLKEPAPTTPKKPAPKELAPTTKGPSTTSDKP-APTTPKETAPTTPKEPAPTTPK 797
Db    628 TTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSS-TTPVQTTTTTAPETTSTEP- 686
Qy    798 KPAPTTPEPPPTTSEVSTEPTTTKEPTTIHKSPDESTTPELSAEPTPKALENSPKEPGVPT 857
Db    687 ---PSSSNTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETT----------STEPPSSST 737
Qy    858 T--KTPAAATKPEMTTTAKDKTTTERDLRTT----PETTTAAPKMTKETATTTEKTTESKIT 911
Db    738 TPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETT 797
Qy    912 ATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTTITTTEIMNKPEETA-KPKDRATN 970
Db    798 STEPPSSSTTSPVQTT--TTTAPETTSTEPPSSSNTPVQTTT--TTAPETTSTEPPSSSTS 853
Qy    971 SKATTEPKPQKPTKAPKKPTSTKKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAML 1030
Db    854 PVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPET- 911
Qy   1031 QTTTTRRNQTPNSKLVEVNPKSEDAGGAEG-ETPHMLLRPHVFMPEVTP---DMDYLPRV- 1085
Db    912 -TSTEP---PSSSTTPVQTTTTTAPETTSTEPPSSSSTTPVQTTTTTVPCDCSLSYIDRVV 967
Qy   1086 ------------PNQGIIINPMLSDETN---------------ICNGKPVDGLTTLRNGTL 1119
Db    968 YPFTTEWWENKRDIIIQSVDSPRRTAFSVSTFDIGTCTAILTCIYTSEGGISNL-NATL 1026
Qy   1120 VAFRGHYFWMLS------PFSPPSPARRTEVWGIPSPIDTVFTRCNCEGKTFFPKDSQ 1172
Db   1027 ------FIGLSDGSSIDLPF-----------YNVPTGLE-IWPEINCEGNK------- 1059
Qy   1173 YWRFTN 1178
Db   1060 -WSYNN 1064


RESULT 6
Q917S1
ID   Q917S1       PRELIMINARY;      PRT;   1049 AA.
AC   Q917S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
```

RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nixon K.A., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003495; AAG22353.1; -.
DR   FlyBase; FBgn0030561; CG5228.
SQ   SEQUENCE 1049 AA;   107278 MW;   954D629E7619671 CRC64;

Query Match   13.0%;   Score 955;   DB 5;   Length 1049;
Best Local Similarity 35.1%;
Matches   376;   Conservative 54;   Mismatches 474;   Indels 168;   Gaps 54;

```
Qy   106 EVHNPTSPPSSKRAPPSGCASQIIKSTIKRSPKPPNKKKTKVIESEEITEVKDNKKNRT 165
             : :    :           ::  :              :  : : :
Db    33 EIGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTE--TAKPTLKPTE--GT  88

Qy   166 KKKPPTPKPPVVDEAGSGLDNGDFKVTTPDTS--TTQHNKVSTSPKITTAKPINPRPSLPP 223
             :    ::                 :  :::       : :      ::
Db    89 TAKPTTLKPT---EGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKP- 144

Qy   224 NSDTSKEFSLTVNKETTVETKETTTTNKQT--STDGKEKTTTSAKET--QSIEKTSAK--D 277
            ::    ::          :  : :::          :   ::: :      : ::::    :
Db   145 --TTLKPTEGTSAKPTTLKPTEG----TSAKPTTLKPTEG  198

Qy   278 LAPTSKVLAKPT--PKAETTTKGPALTTPKEPT---PTTPKEPASTTPKEPTPTTIKSAP 332
                  ||||       :         ::: :       :::           :: ::
Db   199 LKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAK---PTTLKPTE 255

Qy   333 TTPKEPA---PTTTKSAPTTPKEPAPTTPKE---PAPTTTKEPAPTTTKSAP 386
                ||        :|| :      :    :       :    ::
Db   256 GTTAKPTTLNPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAK---P 313

Qy   387 TTPKEPAPTTPKPAPTTPKEPAPTTPKEPT---PTTPKEPAPTTPKEPAPTTPKEPA---P 441
                ||       :      :    :       :::         :
Db   314 TTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLKP 370

Qy   442 TAPKKPAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKSAPTTPKEPAPTTTKS 501
              ||       :         ::: :          ::: :
Db   371 TEGTSAKPTTLKPTEGTTAK---PTTLK---PTEGTSAKPTTLKPTEGTTAK--- 421

Qy   502 APTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTA 561
             :::     ::: :      :    :   ::      :         :
Db   422 -PTEGTTAKPTTLKPTEGTTAK---PTTLK---PTEGTTAKPTTLK-PTEGTTAKPTTLK 473

Qy   562 PKEPAPTTPKEFAPTTPKKLTPTTPEEKLAPTI-TPEKPAPTTPEE---LAPTI--PEEPT 614
```

RESULT   7
O76894
ID   O76894   PRELIMINARY;   PRT;   1795 AA.
AC   O76894;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:5G7.1 PROTEIN.
GN   EG:5G7.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

```
Qy  1018 ELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEG-ETPHMLLRP 1068
              ||  :         |  :        :          :
Db   996 TLKPTERTSAQ--PTTLKPTEGTTAKPTTLKPTKGTSGQANFETKRKERRP 1045

Qy   967 RATNSKRATTPKPQKPPKAPK---KPT--STKKPKTMPRVRKPKTTPT---PRKMTSTMP- 1017
              :  ||    :      :       :       :   :   :
Db   937 -GTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTERTSAQPT 995

Qy   909 KITATTTQVTSTTTQDTTPFKITTLKTTTLLAPKVTTTKK--TITTTEIMNKPEETAKPKD 966
             :|      :        :      :        :
Db   877 PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLKPTEGTSAKPTTLKPTE 936

Qy   866 PEMTTTAKDKTTERDLRTT-------PETTTAAPKMTKEPTATT------TEKTTES 908
              |        :              :           |     :
Db   817 PTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTTAK 876

Qy   825 TIHKSPDESTPELSAKPTP--------KALENSPKPGV--PTTKT---PAATK 865
             :             |:||         :            ::|      |:|
Db   762 TL--KPTEGT---SAKPTTLKPTEGTTAKPTTLKPTERTSAQPTTLK 816

Qy   771 TSDKPAPTTPKE--TAPTTPKEPAPTTPKKPA---PTTPETPPPTTSEVSTPPT-TTKEPT 824
             ||    ||      :     :    :         ::: :    :           ||||
Db   702 TSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTRTTAKPT 761

Qy   715 APTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPK----KPAKRELAPTTTKGPTST 770
                 :    :         ::   |      :         :      |    ||:|
Db   648 EGTTAK--PTTLK---PTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLNPTEG 701

Qy   673 TLKEPAPTTPKKPAPKELAPT-------TTKEPTSTTTSDKPA-------PTTPKGT 714
             ||    :   :     ||         ::|     :    |         :||||
Db   591 TLK---PTEGTSAQPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPT 647

Qy   534 TAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAK----PT 590
              ||      :         ::     :                   :
Db   474 PTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPA--PTTPKEPAPTTPKETAPTTLKPTKGT 533
```

RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbert W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li J., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.;
RL   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003421; AAF45644.1; -.
DR   EMBL; AL031028; CAA19845.2; -.
DR   FlyBase; FBgn0025390; EG:5G07.1.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF01607; Chitin_bind_2; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00494; ChtBD2; 2.
SQ   SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match             11.9%;  Score 868.5;  DB 5;  Length 1795;
Best Local Similarity   28.2%;  Pred. No. 2.4e-49;
Matches  281;  Conservative  134;  Mismatches  460;  Indels  123;  Gaps  29;

Qy    131 STTKRSRKPPPNKKKTKKVIESEEITEVKDNKNRTKKKPTPKPPVVDEAGSGLDNGDFKV 190
              : ||  ||    |  : : :| |||:||:|  |  ||:|
Db    408 ATTREPLNDINKYQYKRTYGYD-------KNDVTEAPEIKSPL----KGLHLSENTV 454

Qy    191 TTPDTSTTQHNKVSISPKITTAKPINPRPSLPPNSDT------------SKETSLTTVNKET 239
          : |:|||                    :::|||:|||               :|:|:|:|::
Db    455 ILPETTTT--TTTTTKPVVLTCPTISPDTTPKPSTTTAVTKSTPKISSTEQHSTTTAKT 512

Qy    240 TVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGP 299
          : | |:||||:|  ||  :|:|||:|: :  |:|::     |:|  |||| |: |||||
Db    513 TTTKRPTTVTEKTSSAIEKPRTTVVTTTTQKRSTTTHNTSPDTKTTIRSTTLSPKTTTP 572

Qy    300 ALTTPKPEPTTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTK 359
          :  |||   || :||  | |||| |  || ::|||||| |  ||||:|  |||
Db    573 STTTPSTTTPST-----TTPSTTTPST-----TTPSTTTTVKVSTHRPRTTSQK 621

Qy    360 -EPAPTTTPKEPAPTTPKEPAPTTPKKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTP 418
           |: |||  |  |||||: |:|:|| |  |||||  |||| || |||||  | |||| |
Db    622 TTTTASTTTKK------TTTSPKTTKTTDIPTSTTRSKLSTTTQKTTTTHKFTAATTSTEKPK 677

Qy    419 TTPKEPAPTTKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKE 478
          ||| | ||| |  | ||| |   ||||||| |: ||| | |: |  |||| |
Db    678 TTTEKTSTVSTTKKSTESSPKPTSSGTGKPTTTPKPSTRTTTPTTKVTTTQITTTPLR 737

Qy    479 PAPTTTKSAPTTTPKEPAPTTTKEPAPTTPKEPSPTTTPKEPAPTTPKEPAPTTPK------ 531
          || ||||:|||| || |||||:|| | |||  |  |||||||| |||| | |||||
Db    738 SSTETTSTQPPTTTPQPTTTTTLTVTPKTVSTTTTTTTTEKPITSSPKPTTTQKTTTSTAPN 797

Qy    532 -KPAPTTPKEPAP--------------TTPKEPAPTTKKPAPTAPKEPAPTTPK-ETAPTT 576
           |||:||:||:|:              |||:|| |||: |||| |:| ||||| | |:||:
Db    798 TTKVAITTQKETTPTQSTSTTIFTTRKTTTTNNPEPTSTEKPTITSTTTPKPSTTTVAS 857

Qy    577 TPKKLTPTPEKLAPTTPEKP------------APTTPEELAPTTPEEPTPTTPEEPAPTTPKA 628
           || | ||:|: |||:||:|            || || | ||| |  ||||| | |||| |
Db    858 STEKTTISSPKPTTEKSTENPTTNSVKTSALTSSTQRAISTTSEPTKTT-QNITTTPKP 916

Qy    629 AAPNTPKEPAPTTTPKEPAPT--TPKEPAPTTPKETAPTTPTKGTAPTTLKEPAPTTPKKPA 686
          :| :|||:| ||:|||: ||:  |||:|:|| |::| ||| | |: || |:||| |||:| |
Db    917 TTLKTSTQEATTSTQKVSTVTITTKKATESSPLTTLSTEEPNTTPKPLRTTTPTTTSVTA 976

Qy    687 PKELAPTTPKEPT--STTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKKPAPTT 744
          |:: | |: :|   |||:||| |||||  || |||:  |||| |   ||| |:|:||
Db    977 TTRITTTTISESSTETTSTQKPKSTTPTSTTTRTTPKVTTVTVVSTQNPTTTSKTTSTVTI- 1035

Qy    745 EPAPTTTPKKPAPKELAP-TTTKGPTSTTSDKPAPTTPKETAPTTPKKEPAPTTPKKPAPTT 803
          :|:||||||:| |:  || |||:| || ||| |||||||:| |||||   ||||||:| ||
Db   1036 ----TTP-NPSPSTQRPTTTTTRQPTSITASTTSIGTTRIPTTTNPQNSTSSTDLTTVTRP 1090

Qy    804 PETPPTTTSEVSTPTT-TKE------------PTTIHKSPDESTPELSAEPTP 843
              ||||| ||:||| |||            ||||:|||| |||| | |:||
Db   1091 PCPDPDSTSDKNTNTPACTQELQQVNLLELQSPQKQEQFTHTRTHFALTGSRNTLGCQEVP 1150

Qy    844 KALENSPKEPGVPTTKPAATKPEMTTTAKDKTTER---DLRTTPETTTAAPKMTKETAT 900
          ||:|:||||||:|||||||||||||||||||||||   |||||||||||||||||||||
Db   1151 DYMDDAPSSAREAESCQATTAKAPTMSTLAAAHLLQKLFHIISTTPPSREHAP--TORPSS 1208

Qy    901 TTEKTTESKITATTTQVTSTTTQDTTPFKITTTLKTT----TLAPKVTTTKKTITTTEIMN 956
          |||:||:|::|||||||||||||:|||||||||||||    ||||||||||||||||||||
Db   1209 QPSSSQSR-GVTIAQMARHNIATSKPFIAHSLRLSIQQLASTQKRSIPPKTLVTHNTTK 1267

Qy    957 KPEETAKPKDRATNSKATTTPKQKPTKA-PKKPTSTKRKRPKMPRVRKPKTTTPTPRKMTST 1015
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   1268 EPED-SEYVDSETSEQYTDEDNEVLDKTQPRAMSSTTVAAVLPAVPSTTTEREPQK-TSS 1325

Qy   1016 MPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSED 1053
          ||:||||||||||||||||||||||||||||||||||
Db   1326 SP--SPT----KATSSTTTQPIETTTGDLEYDSSGSSSD 1357

RESULT 8

ID   O96449
AC   O96449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RACE 1-11;
RA   Goernhardt B.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061185; AAC72308.1; -.
SQ   SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

                                          PRELIMINARY;   PRT;   1489 AA.

Query Match             11.8%;  Score 867;  DB 10;  Length 1489;
Best Local Similarity   32.2%;  Pred. No. 2.5e-49;
Matches  367;  Conservative  52;  Mismatches  503;  Indels  216;  Gaps  51;

Qy    114 PSSKKAPPSGASQTIKSTTKRSPK-----PPNK---KKTKKVIESEEITEVKDNKKNR 164
          ||  :||| |:  ::  ||:|:||     ||||   |:||||||||||||||||:|||
Db    303 PSDETEAPTEGTTYVPREETTAAPSEDTTYAPREVTPYAPTEKPYDVEETTYVTEESTYA 362
```

```
RESULT   9
Q9SPM0
ID   Q9SPM0      PRELIMINARY;   PRT;   1315 AA.
AC   Q9SPM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN.
DE   PEX2.
GN   PEX2.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=POLLEN;
RA   Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA   Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT   "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF152297; AAD59980.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR001998; Xylose_isom.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR   SMART; SM00370; LRR; 3.
SQ   SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;
```

```
Query Match        11.5%;   Score 839.5;   DB 10;   Length 1315;
Best Local Similarity  24.4%;   Pred. No. 1.5e-47;
Matches  265; Conservative 101; Mismatches  49; Indels  227; Gaps  34;
```

```
Qy    73  RCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSK-KAPPSGASQTIKS   131
            :   :|                         : ||| ::   :  : :|
Db   390  QCAPVLARPVEVDCS------------KHVCAGYPTPGGPPSSSVPGKPPSVPGK----   432

Qy   132  TTKRSPKPPNKKKTKVIESEEITEVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVT   191
            : |                              | ::    :
Db   433  --PAAAPAMPIPHTPPDVSPEPLPEPSPVPAPAPMRMPTLRSPPADE-------YIP     480

Qy   192  TPDTSTTQHNKVSTSPKITTAKPINPRPSLPPNSDDSKETSLTVNKETTVETKETTTTNK   251
            ||  |          :|   |    | |
Db   481  TPPVPA--KSPPGTSPPASRGAP--PLQAQPPAASSPPATPV------------------   518

Qy   252  QTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTIKGPALTTPKEPTPTT   311
                  | :|            :     |  :|           :||  : ||
Db   519  --------KSSPPAAVLPPPAKTPSPAPAPV-ASPPPEAPVSSPQPQVKSSPPPAPVA     568

Qy   312  PKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTP-KEPAP--------------   363
            ||  :  |   ||   :| :   ||  : |   :|    :|::
Db   569  SPPPPMKSP--PPPARVASPPPLMKSPPPAPVASPPQPLKSPPPPVIWLSTPSVKSSPP   626

Qy   364  TTPKEPAPTTPKEPAPTTPKSAPTTPKE----PAP--TTP----KKPAPTTPKEAPTTP   413
            : | :   : |    | :| ::     :    :|:  :||      |:|| : ||
Db   627  PVPVASPPPPVKSPPPLAPVSSPSPPPVKLPPLPAPGKSTPPPEEEKRTPPPPVKSSPPPE 686

Qy   414  KE-PTPTTPKEPAPTTPKEPAPTTPKEPAPTTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPS 472
            | :| :| : ||  : | :  :|  ::   | | | :||  : | :   :| :  : | :
Db   687  KSLPPPTLTTSPPPQEKPTPPSTPSKPPPPSPVETLPPPSKSSPPEEPVSSPPQAPKSSS   746

Qy   473  PTTPKEPAPTTPKSAPTTTKEPAPTTPKEPSPTTPKEPAPTTPKK               532
            |  | :   :| | :|     :| : | :  : |   : | | |
Db   747  P-----PAPVS--SPPLKSPPPVPESSPPPTPKSSPPLAPVSSPPAPVS           799

Qy   533  PAPTTPKEPAPTTPKEPAPTTPKKPAPTTPKETAPTTPKETAPTTPKKLPTTPKLAPT    592
            || | | :   :| | :   :| :  |  : | :  :| : | : :   || :|| | :|
Db   800  SPPPTPKSSPPLAPVSSPPQVEKTSPPPAPVSSPPPTPKSSPPLAPVSSPPQVEKTSPPP  859
```

```
Qy   165  TKKKPTPKPPPVVDEAGSGLDNGDFKVT------TPDTSTTQHNKVSTSP-KITTAKPINPRP  219
            |||
Db   363  PTKSETNAPTERMHYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAPTEETP    422

Qy   220  SLPPNSDISKETSLTVNKETTVETKETTI---TNKQTSTDGKEKTTSAKETQSIEKTSAK    276
            :::|| :   |::|:  |  :|      :|
Db   423  YEP------TEETTYTPTEETTYAPTEETTYAPTEETTYAPTEKTTYAPTEETPYEPTEET   477

Qy   277  DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAPTTPKEPATTIKSAPTTPK    336
            :|       || : | ::::        |:||: |  ||| | : ||::
Db   478  TYAPTEETTYAPT---EETTYASTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEET    533

Qy   337  EPAPT-TTKSAPTTPKEPAPT---TTKEPAPTTYAPTEETTTKSAPTTPKEPAPTTPKEPAPT 387
            :|||  || :|| ||    | :   :||:||: | ||    :| :|| ||  | :|
Db   478  APTDKKPAPTTPKEPAPTTYEPTEETTYAPTEETTYASTEETTYAPTEETTYAPTEETP    533

Qy   388  ----TPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEP-----------------AP-  426
            :   :| :|| | | : || | : || | : |: |: | |                 ||
Db   593  PIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEE  652

Qy   427  TTKEPAPTTPKEP----APTAPKKPAPTTPKEPAPTTPKEPAPT----TTKEPSPTTTPKE  478
            :| :|||: | :|    ||| | | || | : ||| | : |::     :| :| :|| | :
Db   653  TTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYE  712

Qy   479  P------APT-TIKSAPT-----------------APT-TIKSAPT---TPKEPSPT     512
            |      ||  || :|||                 ||  || :|||    | :| :|
Db   713  PTEETTYAPTEETTYAPIEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYE    772

Qy   513  TTKEP------------APTTPKEPAPTTPKKPAPTTPKEPAPT---TPKEPAPTTTKKPAPTAPKE 564
            :| :|            || | :| |:|  | || | : || |    | :|| :| ||| |:|
Db   773  TTYEPTGETTYAPTEETTYAPTEETTYAPTEEETTYAPTEETP---YEPTEETTYAPTEETPYE    832
```

```
Qy    593 TPEKPAPTTPEELAPTTPEEPAPTTPTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKE 652
              |||  :| ||||  :||  ||:
Db    860 APVSSPPLEPKPSSPPSVSSPPTVKSSPPPAPLSSPPMTPKSSSPPAHVSSPPEAEKS 919

Qy    653 PAPTTPKETAPTTPKGTA-----------PTTLKEPAPTTPKPAPKELAPTTKEPTST 701
              :| |                      |:|
Db    920 SPPLAPISSPPSSEPKSPSSPPMVEKTSPPPATVSSSPPPTPKSSPPP---APVSSPPPVVK 976

Qy    702 TSDKPAPT----------TPKGTAPTTP--------KEPAPTTPKEPAPTTPK--GTAPTTLKEP 746
                                 ||||         |:|
Db    977 SSPPPAPVSSPPPTPKPLPPPAPVSSPPPPTPVVKSSPPPTPKGTAPTTLKEP 1036

Qy    747 APTTPKKPAPKELAPTTTPKGPTSTTSDKPAPTT--PKETAPTTPK----EPAPTTPKKPA 800
              |:|        ||  |||                |:|
Db   1037 PPPXKPLPPP---APVSSPPPVVKSSPPPAPVSLPPPPIRKPSPPRTRVSSSPRPVVKCCPP 1093

Qy    801 PTTPEPTPPPTTTSEVSTPTTPTTTKEPTTIHKSPDESTPELSRAEPTTKRALENSPKEPGVPTTKYT 860
              ||:|                |:|
Db   1094 PTLVSSPPPAPKSLPPPPTPVVSSP----PPEVKSSPPPTPVSSPPPSPPAPKSSPPPT 1143

Qy    861 PAATKPEMTTTAAKDKTTERDLRTTP-----ETTTAAPKMTKETATTTEKTTESKITATTT 915
              |:|                            |:|
Db   1144 PVSSPPP------ELKSSPPPAPVSSPPSSAPKSSPPPAPVSLPPPPEVKSSPPPA 1191

Qy    916 QVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKITTTTTELMNKPEETAKPKDRATNSKATT 975
              |:|
Db   1192 PISSPPPPAKSP-------------PPPA 1225

Qy    976 PKPQKPTTKAPKKPPTSTKKPKTMPRVRKPKTTPTPTPRKMTSTMPELNPTSRIAEAMLQTTTTR 1035
              |:|
Db   1226 SSPPPPMKSPPPPAPISSPPPPAP--VKPSLPPPAPVSSPPPAV----------TSA 1270

Qy   1036 PNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPM 1095
              |:|
Db   1271 P----------PKKEEDSTA 1299

Qy   1096 LSDE 1099
              :::
Db   1300 MANK 1303
```

RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaughan M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Leimbach D.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RL   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U46675; AAB52641.1; -.
SQ   SEQUENCE 1274 AA; 138065 MW; 07F6B0D292C1799F CRC64;

```
Query Match     11.4%;     Score 835;   DB 5;   Length 1274;
Best Local Similarity   27.5%;   Pred. No. 2.8e-47;
Matches 310;   Conservative 111;   Mismatches 446;   Indels 260;   Gaps 60;

Qy     95 KCCPDYESFCAEVHNPTSPPSSKKAP--------------------PPSGASQTIKSTTKR 135
              |:|                                        |:|
Db    189 KAAPSKE-------HDPIVPPTPIKNPAKKWKPPWEDDEVPTEEIKEPEPATRKVPALKKK 242

Qy    136 SP-----KP-PNKKKTKVIESEEITEVKDNKKNRTKK-KP------TP------KPPVVDEA 179
              |:|       ||||                       |:|
Db    243 EPSTSVKPVSDPSPTKVPVKKEPEVPPTPIKNPTKKWKPWEDEDTPVEEVKEPPVPEKK 302

Qy    180 GSGLDNGD-----FKVTTPDISTSTQHNKVSTSPKITTAKPI-NPRPSLPPNSDTSKETSLT 234
              |:|          ||||
Db    303 APVLKKKDPAPAAKARDPSPSKAAAPKKVEPSSPVVPPTPVKNPVKKVKPPWEVDDEPAEE 362

Qy    235 VNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIERTSARDLAPTSKVLAKPTPKAE- 293
              |:|
Db    363 VKKPSAPEKKTPVLKRKEPEPSSTTPSSDPSPKKAAPAVKPRDSSPKKATPLQADPKAQE 422

Qy    294 --TTTKGPA------------------LTTTKEPTPTTP-----KEPASTTPKEP------TP 325
                 |:|                    ||||            |:|
Db    423 VPPTPVKNPVKVKYKPPWEVDDEDPVEEVKQPEAPAKTPVLKRKEPAAAKDTAKPATSKTP 482

Qy    326 -TTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPA-------------------PTT 365
              |:|                                   |:|
Db    483 ETPEKKDPVKPRDSSPKKVAAKPDSAQAPA-TPVKNPVKKRPPWEDDETPADDVSKPTD 541

Qy    366 PKEPAPTTPKEPAPT----------TTKSAPTTPKEP-----APTTPKPAPTTPKE-----PA 409
              ||||                     |:|
Db    542 AKKTPSLAKKDPAPAKESLKPKADTKAPAKPRDPSPKKVAAPTAPAEKKTPVLAKKEPAGPA 601

Qy    410 PTTTPKEPTPTTPKEPAPTTKEPAPTTPK-EPAPTAPKKPAP------TTTPKEPAPTTPKE 462
              |:|                         |:|
Db    602 DSKTKEPEKSKPRDPSPKKAVPAKPVPKTEVAPAAVKKPEPISKPKDTAPKKAEPNSPVV 661

Qy    463 PAPTTTKEP----SPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTP-KEPSPTTTKEP 517
              ||||        |:|
Db    662 P-PTPVKNPVKKWKPPWEDDDAPAKPVSLPPEPEKK-TPVLAKKAPTKPDSEAAADPVSGP 719

Qy    518 APTTPK--EPAPTTPKKPAP--TTPKEPAPTTPKKPAPTTS-PKEPAPTTPKE 572
              |:|                  ||||
Db    720 SSKDPKLAKKAPVKPRDSSPKKVAAKPDSPMKAVPIKPAPKT--EVPPAVVKKPEPVAKSRDPSPKKAK- 776

Qy    573 TAPTTPKKLTPTTP-----EKLAPTTPEKPAPTTPEELAPTTPEEP----TPTTPEEPA 622
              |:|                ||||
Db    777 AEPNSP--VVPPTPVKNPVKKWKPPWEDDDAPAEPVNVPEPEKKTPVLAKKTPVKRDPS 834

Qy    623 PTTPKAAAPNTPKEPAPTTPKEPAPTT-PKEPAPTTPKETAPTTPKGTAPTTLKEPAPTT 681
              |:|                        ||||
Db    835 PKKAVPAKPSRTDAPVSVKKPEPVSKKPEPSKKAEPNSPVVP----PTPVKNPVKKW 890

Qy    682 PKKPAPKELAPT-TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKGTAP 740
              |:|                      ||||
Db    891 -KPPWEDDDEPTEEVKKPSE--PEKKTPVLAK-KEPEKPKD-APKVAAKPRDSPKKAVP 945
```

```
Qy   741 TTLKEPAPT---------TPKKPAP---------KELAPTTTKGPTS----------TTSDKP-----AP 777
             |||||                          :|| | |||              ||:|         |
Db   946 E--KEPAKVAAKPRDLSPKKAIPIPANTQEAPPTPVKNPVKKWKPPWEDDDEPAEPVSAP 1003

Qy   778 TTPKET-------APTTPKEPAPTTPKKPAPTTPETPPPTTPSEVSTPTTTKEPTIIHKSP- 830
                       :|:|  |||| ||| ||  |||   ||  |:| |||||| |  ||
Db  1004 EPEKKTPVLAKKAPAKPRDP---SPKKAAPVAAK-PDDKIPEV-PDTPVKNPVKWKPPW 1058

Qy   831 ---DESTPELSA-EP---TPKALENSPKEPGV-------------PTTKTPAAT------ 864
                 |  |      :|                         || || |||
Db  1059 EDDDEPSEPVSAPEPEKKTPVLAKKAPTKPATKPDSEAAADPVSGPTSKDPKLSKKAPVE 1118

Qy   865 KPEMTTAKDKTTTERDLRTTPETT--TAAPKMTKETATTTEKTTESKITATTTQVTSTTT 922
             || ||  ||      |  |||     |||| |                  |: |||
Db  1119 KPKPTTDPKDDKLKPSPAKKPEKAPEPAAPKKWKPVWDDDPDEPEADFTVPAPSKKPDTE 1178

Qy   923 QDTTPFKITTLKTTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKATTPKPQKPT 982
             |                  |                                    ||
Db  1179 DPADPLG-------GPKTKDPK--------------LNKKAPAEKPTEK------ 1213

Qy   983 KAPKKPTSTKKPKTIMPRVRKPK------------TTPTPRKMTSTMP 1017
             | |||                           ||| ||  |:
Db  1214 KEPRKPTEPPKP-AAPKKWKPPWEDDPDEPEEADFTMPAPKKPDTEDP 1259
```

```
RESULT 11
Q9LIE8
ID   Q9LIE8;     PRELIMINARY;     PRT;   1480 AA.
AC   Q9LIE8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP001306; BAB03062.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   1480 AA;   147153 MW;   D1AC0C79F155E732 CRC64;
```

```
Query Match     11.3%;   Score 830;   DB 10;   Length 1480;
Best Local Similarity   29.0%;   Pred. No. 7e-47;
Matches  321;   Conservative  62;   Mismatches 539;   Indels  186;   Gaps  48;
```

```
Qy   110 PTSPPSSKKAPPSGGASQTIKSTTKR-------SPKPENKKKTKVIESEEITEVKDNKNRTK 166
Db    46 PPRPPAVKPHPHPK--PPIKPPPEKRHHPHKPPTVKPHP---HPKPPTKPHPHPKPPTK 100

Qy   167 KKPTPKPPVVDEAGSGLDNGDFKVTTPDTTSTTQHKVSVSTSPKITTTAKPINPRPSLP--- 222
Db   101 PHPHPKPPPI-------KPPPHPKPRPHPKPPNVKP-HPHPKPRPTKPH 140

Qy   223 PNSDTSKETSLITVNKETTVETKETTITTNKQTSTDGKEKTTSAKETQSI 270
Db   141 PHPKPPTKHHPHPKPPTIKPPPKPPPSVKPPSTPKPPTTNPPPSTPQPPHKPPPSCTPTP 200
```

```
RESULT 12
P70670
ID   P70670;     PRELIMINARY;     PRT;   2187 AA.
AC   P70670;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
```

```
Qy   271 EKTSAKDLAPTSKV--LAKPTPKAETTTKGPALTTPKEPTP--TTP--KEPASTTKEPT 324
Db   201 PVASPPMATPPTQMPPIATP-PIAKSPVATPPIATPPTATPPITTPPPVATPPITTPIAN 259

Qy   325 PTTIKSAPTTPKEPAPTTT-------KSAPTTPKEPAPTTT-----------KEPAPTT 365
Db   260 PPIIMPPIATPPVAAPPITNPPISKPPVTTP----PTTTPPIAKPPIATPPISTPPAATP 315

Qy   366 PKEPAPTTTKEPA----------PTTTKSAPTTP-KEPAPTTPKKPAP--TTPKEPAPTTPKE 415
Db   316 PAATPPIITTLPPAKPPVAISPIVT--PPVTPIAQPPVATPPTATPPVATPPIATPPTSKS 373

Qy   416 PPPTTPKEPAPTTPKEPAPTTP-KEPAPTAPKKPAPTTPKEPAPTTPKEPAPTT----KE 470
Db   374 PISTPPISESPVATPPTATPSIKTPPPAKPPVATPPIAKSPIATPPTATPPVATPPIEKP 433

Qy   471 PSPTTPKEPAPTTTKSAPTTTKEPAPTTPKEP--SPTTTKEPAPTTPKEPAP- 527
Db   434 PVATPPTTP-PIAT--PPVAKPPVETPPIATPPTAKPPISTPPISKPPVATPPAATPPI 490

Qy   528 --TTPKKPAPTTPKEPAPTTPKEP--APTTTKKPAPTAP--KEPAPTTPKEPAPTT--- 577
Db   491 TTPPTPVKPPVATPPLAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVATPPI 550

Qy   578 --PKKLIP--TTPEKLAPTTPEKPAPTTPEELAPTT-PEEPPPTTPEEPAPTTPKAAAPN 632
Db   551 AKPPVVTPPTTTPPTTATPPTPPVATPPVATPPIATPPTATPPIATPPIISKSPVATPPAATPPI 610

Qy   633 T----PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKETAP--TTPKGTAPTT---LKEPAPTTPKK 684
Db   611 TPPPPAKPPVATPPIATPPIAKPPVATPPTATPPIATPPVAKPPVATPPIKTPPPAKPPV 670

Qy   685 PAPKELAPTTTKEPTSTTSDKPAPTTPKGTAP--TTPKEPAPTTPKEPAPTTPKGTAPTTL 743
Db   671 AIPPIATPPVAKPPVAT---PPTATPPIATPPVTPPTATPPVATPPIAKPPPTI 727

Qy   744 KEPAPTTPKKPAPKELAPTTTKGPTST--TSDKPAPTTPKETAPT-TPKEPAPTTPKKPA 800
Db   728 -PPTATPPVAMPPIATPPTAKPPIATPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPV 785

Qy   801 PTTP-------ETPPPTTSEVSTPTTTKEPTTIH--KSPDESTPELSAEPTPKALENSPKE 852
Db   786 ATPPIATPPVATPPIAKPPVATPPTATPPIATPPVAKPPVATPPIATPPIATPPAKPPIL 845

Qy   853 PGVPTTKTPAATKPEMI--TTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKIT 911
Db   846 SKPPVATPPAATPPIITTPPPAKPPVAIPPIATPPPTIPPVATPPIATSPVAKP 903

Qy   912 ATTTQVTSTTTIQDTTPFKITTLKTTTLAPKVTTTKKI----ITTTEIMNKPEETAKPKDR 967
Db   904 PVAIPIKTPPPAKPPVAIPPIATPPVAKPPVATPPTATPPIATSPIATPPVVT---PPT 960

Qy   968 ATNSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRKPKTPTPRKMTSTMPELNP------ 1021
Db   961 ATSPVATPPIAKPPTTTP--PIAT--PPVAMPPIATP---PTAKPPVATPPIANPPVEKPP 1014

Qy  1022 --TSRIAE--------AMLQTTTRPNQTPNSKL--VEVNPKSEDAGGAEGETPHMLLRPH 1069
Db  1015 VATPPIAKPPTVLPPIAKPPVETSPTATPPTATPPVAIPP------------------VVKPP 1059

Qy  1070 VFMPEVTPDMDYLPRVPENQGIIINPMLS 1097
Db  1060 VAIPPITKPPVATPPVTNPPTAMPPIVT 1087
```

```
DE   NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE   MUSCLE-SPECIFIC FORM GP220).
GN   NACA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96312450; PubMed=8698236;
RA   Yotov W.V., St-Arnaud R.;
RT   "Differential splicing in of a proline-rich exon converts alphaNAC
RT   into a muscle-specific transcription factor.";
RL   Genes Dev. 10:1763-1772(1996).
DR   EMBL; U48364; AAB18734.1; -.
DR   EMBL; U48363; AAB18732.1; -.
DR   MGD; MGI:106095; Naca.
DR   InterPro; IPR002715; NAC.
DR   InterPro; IPR003037; TS-N.
DR   Pfam; PF01849; NAC; 1.
DR   Pfam; PF02094; TS-N; 1.
SQ   SEQUENCE 2187 AA; 220599 MW; 003646AA864DEBFD CRC64;

Query Match             11.3%;  Score 828.5;  DB 11;  Length 2187;
Best Local Similarity   28.4%;  Pred. No. 1.3e-46;
Matches  313;  Conservative  126;  Mismatches  436;  Indels  229;  Gaps  52;

Qy    109 NPTSPPSSKKAPPS-GASQTIKS------TTKRSPKP--PNKKKTKKVIESEEIT 155
          ::: ||||||| : |  |  ||          ||:||:|  |:||||||||||:|||
Db    746 SPT--PPSSKGAPVPSTGAPPSPRKGAPIVPTESSISSKQVPAEILPSPQKTPEVTASRLIS 804

Qy    156 EVKDNKKNRTKKKPPTPKPPVVDEAGSGLDNGDFKVTTPDTSTIQHNK-----VSTSPKIT 210
          | ||||||||||||| :||    |          | :  :  |         ||:| :||
Db    805 AVGSPKVDPIMSDVTPTSPKKTSA------TAVPKDTSATLSLKSVPAVTSLSPPKA 855

Qy    211 TAKPINPRPSLPPNSDTSKETSL-----TVNKET------TVETKETTTNKQTSTDGKEKTT 262
          |  | | |  |  |:  ||: :|      ||:|:       ||: |: : |: |  |: ||
Db    856 PVAPSNEATIVPTEIPTSLKNALAAAATPKETLATSIPKVTSPSPQKTPKSVSLKGAPAMT 915

Qy    263 SAKETQSIEKTSARDLAPTSKVLAKPTPKAETTTKGPALITPKEPTPTTPKEPASTTPKE 322
          | |:   |:::    |  ||:  : | ||   | :    :||:  | :||||:  :||
Db    916 SKKAT---EIAASKDVSPSQ--FPKEVPLLQHV----PPTSPSKSPVSDTLSGALTSPPK 967

Qy    323 PTPTTIKSAPTTPKE-PAPTTTKSAPTTPK------EPAPTTTKEPAPTTPKEPAPTT 372
          | | | ||:   :|  || :| |:  :  |         || | |  |  |  | |
Db    968 GPPATLAETPTYPKKSPRPAASKTPATSPEGVTAVPLEIPCSKKAPKTAAPKESSAT 1027

Qy    373 TTKEPAPTT--TKSAPT------TPKEPAPTTPKK-PAPTTPKEP 408
          ::| || ||  ||:|:        ||: || |||  || ||||:|
Db   1028 SSSKRAPKTAVSKEIPSKGVTAVPLEISLPLKETSKSATPGEKSASSPKRSPKTPAGPKE- 1086

Qy    409 APTTTKEPPTTPKEPAPTTPKEPAPTAPK--KEPAPTT---TTPKE 462
          :| :|: | : || || | |: | :|:      | |||   ::||
Db   1087 -TPPGGVTAVPPEISLPPKETPQNATPNESLAASSQKSPSKTSVPKETPPGGVTAMPLE 1144

Qy    463 PAPTTTKEPSPTTPK-----EPAPTTTKEPAPTTPK------KSAPTTPK-----E 508
          || :| | |   :|       || | |  | :: |         ||| |     |
Db   1145 IPSAPQKAPKTAVPKQIPPEDAVTILAGSPLSPKKASKTAAPKEAPATPSVGVIAVSGE 1204

Qy    509 PSPTTTKEPAPTTPKEPAPTTPKPAPTTP--PKE------RAPTTPKEPAP--- 551
          : |: | ||| | |: :| :| | |: |    |:       |||||||:|
Db   1205 ISPSPKKTSKTAAPKENSATLPKRSPKTAAPKETPAISSEGVTAVPSEISPSPPTPASK 1264

Qy    552 ----TTTKPAPTAPKEPAPTTPKETAPTTPKKLTPTTTPEKLAPTTPEK-PAPTTPEELA 606
              :||| || || : | ||| : ||         | :|  | |:|| | | |  |
Db   1265 GVPVTLTPKGAPNLAE-SPASPKKVPKTVAAPEETSTTP-----SPQKIPKVAGPKEAS 1317

Qy    607 PTTPEEPTPTT------PEEPAPTTPKAAANTPKEPAPTTPKE-PAPTTTKEPAPTT 657
          | ||  | |||       | | || |     ||||| :|||  || :| |  | ||
Db   1318 ATPPSKKTPKTAVPKETSAPSEGVTAVPLEIPSSPRKAPKTAAPKETPAPS--PEGATTA 1375

Qy    658 PKETAPTTPKGTAPTTLKEPAPTTP------KKPAPKELAPTTPKE 697
```

```
Db   1376 PVQIPPSPRKGSKAGSKE-TPTTPSPEGVTAAPLEIPISSKKTSKMASPKETLVTPSSK 1434
Qy    698 PTS-----TTPSDKPATTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPAPTT 750
          |:         ||||| ||||| || ||||| :|||  || ||||||| :||||||||
Db   1435 KLSQTVGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLTPSPK-DAPTTLAE-SPSS 1492

Qy    751 PKKPAPKELAPTTTKGPTSTTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETP--- 807
          ||| ||||:  : :   :: |:  :  | | |:  :    || |  : || ||| |
Db   1493 PKK-APKTAAPPSER-VTTVPPEKPA-TPQKASGTTIASKVPVPAETTQEVAVSSRETPVTP 1549

Qy    808 --PPTTSEVSTPTTTK------EPTTIHKSPDESTPELSAEPTPKALENSPKE-PGVPTT 858
            || :   | :: |          :| | |:  ||  |:| |   | :||| | |
Db   1550 AVPPVRKNPSSHKKTSKTIELKEAPATLPPSPTPKSPKIPSSKKAPRT--SAPKEFPASPSI 1607

Qy    859 KTPAATKPEMTTTAAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVT 918
          | |   :   :: |  | :| |   ::| |   || ||:| | ::||:|| ||| || |
Db   1608 K-PVTT--SLAQTAPPSLQKAPSTTIPKENLAAPAV----LPVSSKSPAAPARASASLSP 1660

Qy    919 STTTQDTTPKITTLKTTTLAPKVTTTKKTITTTELMNKPEETAKPKDRATNSKATTTPKP 978
          :|| ||| ||   : |   |||:| :| || | | : ||  :| ||   |  ||| |||
Db   1661 ATAAPQTAPKEATTIPSCKAAAATEIPIETSTAPSLEGAPKETSE----TSVSKVLMSSP 1716

Qy    979 QKPTKAPKKPTSTKKPKTMPRVRPKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQ 1038
          ||| ||  :    :: :   |      |:   | |:|       |: :  |  |  |
Db   1717 -----PKKASSSKRASTLP------ATTLPSLKEASVLS------PTA 1747

Qy   1039 TPNSKLVEVNPKSEDAGGAEGETP 1062
          | :||  :  ||| :| ||:| |
Db   1748 TSSGKDSHISPVS-DACSTGTTTP 1770

RESULT 13
Q41805

ID   Q41805       PRELIMINARY;        PRT;      1188 AA.
AC   Q41805;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   Rubinstein A.L.; Broadwater A.H.; Lowrey K.; Bedinger P.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z34465; CAA84230.1; -.
DR   Mendel; 14346; Zeama; 2368;14346.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   SMART; SM00370; LRR; 4.
KW   Signal.
FT   SIGNAL        1     27       POTENTIAL.
SQ   SEQUENCE 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

Query Match             10.9%;  Score 800;  DB 10;  Length 1188;
Best Local Similarity   28.7%;  Pred. No. 5.4e-45;
Matches  246;  Conservative  64;  Mismatches  401;  Indels  146;  Gaps  36;

Qy     82 RECDCDDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAPPSGASQTIKSTTKRSPKPPN 141
          |  |    |    :    |     |   :|| |||||||||| || ||||||||||| |
Db    397 RPVDCS------KHVCAGYPTPGGGPPSSPVPGKRPAASAPMPS----PHTPPD 439

Qy    142 KKKTKKVIESEEITEVKDNKKNRTKKKPPTPKPPVVDEAGSGLDNGDFKVTTPDTSTIQHN 201
          | :||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    440 -----VSPEPLPPSPVPAPAPMEPMPTPHSPPAD------DYVPPTPPVPGKSPP 483
```

```
Qy   202 KVSTSPKITTAKPINPRPSL-----PPNSDTSKEPSLTVNKETTVETKETTTNKQTSTDG 257
             |||                  |:|    |||
Db   484 ATSPSPQVQPPAASTPPPSLVKLSPPQAPVG------------------------SPPP 518

Qy   258 KEKTTSAKETQSIEKTSAKDLAPTSKV-----LAKPTPKAETTIKGPALTTPKEPTPTTP 312
                                           ||||
Db   519 PVKTTSPPAPIG-----SPSPPPVSVSPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVAS 574

Qy   313 KEPASTTPKEPTPTTPTIKSAPTTPKEPAPTTP--KEPAPTTTKEPAPTTPKEPA 370
             :|              :|:               ||
Db   575 PPPPVKSP---PPPLVASPPPPVKSPPPPTPVASPPPPPAPVASS 632

Qy   371 PTTTKEPAPTTTKSAPTTP-KEPAPTTPKKPAPTTPKEPAPTT------PKE---PTPTTP 420
             |||                 ||||
Db   633 PPPMKSPPPTPVSSPPPPEKSPPPPAKSTPPEEYPTPPTSVKSSPPPEKSLPPPTL 692

Qy   421 PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPA 480
             :|||:|:|  ||:|  :|:                     :|
Db   693 IPSPPPQEKPTPPSTPSKP-PSSPEKPSP--PKEPVSSPPPQTP---KSSPPPAPVSSSPP 745

Qy   481 PTTTKSAPTTTKEPSPTTTKEPAPTTPKEPAPTTPKE 540
             |:|   |||
Db   746 PTPVSSPPPALAPVSSPPPSVKSSP---PPAPLSSPPPAPQVKSSPPPPVQVSSP--PPAPKS 800

Qy   541 PAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAP------TTP---KKLPTTPEKLAPTT 592
             ||||                            ||:
Db   801 SPPLAPVSSPPPQVEKTSPPPAPLSSPPPLAPKSSPPHVVVSSPPPPVVKSSPPPAPVSSSPPL 860

Qy   593 TPEKPAP-----TTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE-PAPTTPKEPA 646
             |:|            :|:|:|||          ||:|
Db   861 TPKPASPPAHVSSPPPEVVKPSTP--PAPTTVISP-PSEEKSSPPPTPVSLPPPIVKSSPP 917

Qy   647 PTTPKEPAPTTPKETAPTTPKGTAPTTTLKE--PA----PTTPKK---PAPKELAPTTP 695
             |||:|:||             :|:|        ||      ||:||
Db   918 PAMVSSP-PMTPKSSPPPVVVSSPPPTVKSSPPPATPKSSPPPAPVNLPPPEV 976

Qy   696 KE---PTTSTDKPAPTTPKGTAPTTTP-KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 751
             ||      |:|  ||||                |:|
Db   977 KSSPPPTPVSSPPPA---PKSSPPPAPMSSPPPPEVKSPPPPAPVSSSPPPAP 1033

Qy   752 KKPAPKELAPTTTRGPTSTTSDKPAPT-TPKETAPTTPKEPAPTTPKKPAPTTPETPPPT 810
               |||                :|  ||:|       :|:||         |:|
Db  1034 VSSPP----PPVKSPPPPAPVSSSPPPPVKSPPPPAPISSPPPVKSPPPPAPVS--SPPPP 1088

Qy   811 TSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTT 870
             |:|                 :|||                :|:||
Db  1089 VKSPPPPAPVSSPPPPIKSPPPPAPVSSPPPAP-----VKPPSLP-PPAPVSSPPPVVT 1141

Qy   871 TAKDKTTERDLRTTPET 887
             |:|
Db  1142 PAPPKKEEQSLPPPAES 1158
```

RESULT 14
Q9VEL9

```
ID   Q9VEL9      PRELIMINARY;     PRT;   2112 AA.
AC   Q9VEL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG4090 PROTEIN.
GN   CG4090.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
```

RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).

```
DR   EMBL; AE003716; AAF54402.1; -.
DR   FlyBase; FBgn0038492; CG4090.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF01607; Chitin_bind_2; 10.
DR   SMART; SM00494; ChtBD2; 11.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS00022; EGF_1; 1.
SQ   SEQUENCE 2112 AA;  219547 MW;  B91018E5551A5D36 CRC64;
```

Query Match   10.38%; Score 757.5; DB 5; Length 2112;
Best Local Similarity   26.7%;
Matches 296; Conservative 104; Mismatches 348; Indels 361; Gaps 47;

```
Qy   123 SGASQTIKSTTKRSRPKPPNKKKTKKVIESEEITEVKDNKK----------------- 162
             |||                 :|| ||
Db   955 SSSNQTSSSTTQKPFRPAEK------CESEE-TFLADNENCSKFYRCGVDNGKGGFTKVSF 1007

Qy   163 ------------------NRTKKKPPTPKPPVVDEAGSGLDNGDFKVTTPDPTSTTQHNKVS 204
                               :||| |:||         | |:|::| ||
Db  1008 TCPPNTLWDPEANSCNHPDQIQTKPPLCKCKKVVSQGGSSSN-----------STSNSSSSSSNNSGS 1061

Qy   205 TSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTV--ETKETTTTNKQTS----TDG 257
             || |                             :|||          ::|:|
Db  1062 SSNSGGSSSS-------SSSSNGGSSSNTGSSNGASSGGGSSNQGSSNSGSSSGSNSSG 1115

Qy   258 KEKTTSAKETQSI------EKTSAKDLAPTSKVLAKPTPKAETTIKGP----------- 299
             :||                                              ::
Db  1116 NQSTSSTSSSSSSSSSNNNQGSSSSSSSSSSSTSSKPNPSETCKVNGQFIGDRSDCAKFY 1175

Qy   300 -----------------AITTTKEPTPTTPKEPASTT 319
                              |:|||:||
Db  1176 RCVDNDRGGFNMVFSCGPGTVWDAQMQACNIHAWAVKECGGIAPPTTSTTPTTSR-PTTAS 1234

Qy   320 PKEPTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT------ 373
             |:|                           |:|        |||
Db  1235 TSRPSDQTTSRPTGP---PTTARPVTARPTTSSPPTTASSSQTTSPVTQAPNTDGKCRS 1290
```

```
Qy   374 --------------------------TKEP---------APT----------------TTKSAPTTP 389
                                   ::|  :||             |||
Db  1291 EGFMADPNNCSKFYRCVRNNKGGFTSIPFQCGAGTVVWDQDLQTCNHNFNNCSTGTESTTP 1350

Qy   390 KEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKPAP 449
         :|:|:  |||  ||||      |||      ||
Db  1351 KPPC--EPATNGTTATSTSSTTTP--PPTTTDLPPTSTTGLP-PTTTTELPPT 1398

Qy   450 TTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKEPAPTTPKSAPTTPK-- 507
         |||  ||||    |||      |||  ||||
Db  1399 TTTDLPPTTTTRLPTTTTTLPPTTTTTGLPPTTTTTGAQPTTTTLSSETETSTVTTSPEST 1458

Qy   508 -EPSPTTTKEPAP--------------------TTPK------- 523
          |:|  ||||                      |||
Db  1459 TQPPSTTTMKPLPAGTECTGEGYMADPEDCRKYYRCINAGASYRKYNFTCPKGTGWNEEV 1518

Qy   524 ------EPAP---TTPKPAPTTP--------------KEPAPTTPKE-PAPTTKKPAPTAPKEPAPTT 569
               |:||   ||| |||||              |:|||||||  |||| : ||||  ||:|||||
Db  1519 QTCDYVENIPRCSKLPAEPITTTPSEESKDPGSTTPQSTDEPTTVTKPI-TKPTE-EPST 1576

Qy   570 PKETAPTTPKKLTPTTPEKLAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAA 629
         |:|||||||  |||| :|| |||||  ||||    |:||||||  |||||
Db  1577 ERPQKPTTQYPEKPTTTEE--PEKPQKPTI-TTEYPQKPTTTEEPTPEKPQKPI-TTEYPQ 1632

Qy   630 APNTPKEPAPTTPKEPAPTT----PKEPAPTTPKETA--------PTTPK--GTAPTTTIKEPAPTT 681
         |  ||||||||||  |||||    ||||||||||            |||:|  |:| ||| ||||||
Db  1633 KPTTTEEPTPEKPQKPTTTEYPQKPTTEEPTTTSIPGYNPTTTSVPGYNPTTTPIPVETT 1692

Qy   682 PKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAP 740
         |: |:| |||||||||||| :||||||||||||||||||  ||||||| |||||||  ||
Db  1693 TSTPGYK---PTTTTGEPTITTTT-LPSTTTDAIQEPTTTSKKPEEPTTTTTTESPESSTPEGSVT 1748

Qy   741 TTLKEPAPTTPKK------PAPKELA----PTTTKGPTSTTSDKPAPTTPKETAPTTPK 789
         |||||||||||       |||||||    |||| |||||||||||||||||  ||||
Db  1749 TLQPEPQPNYNCSSEGFPPDEDCSRYYRCVDAAKNGKXQVYAFKCGKGTVWDTSTETCN 1808

Qy   790 EPAPTTPKKPAPTTPETPPPTTSEVSTPTTTIHKSPDESTPELSAEPTTPKALENS 849
         ::|: || |||||||| |||||||||   ||:||
Db  1809 YADQVSGN-----------CSSGQTTTPGTTPEPGTTPESTTSSGKP---ETTSKAPEN- 1852

Qy   850 PKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESK 909
         |:|| |||||| ||||||||||||||||||:||||||||||  |||||||||||||||
Db  1853 ------TTTWA---PETTT------SSPETTTT---VASETTTTTSGT----------------- 1883

Qy   910 ITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAT 969
         |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  1884 ---TTTATPETTTKPPKP----------ETTTIAGEETSTSKSPTTTE------------- 1918

Qy   970 NSKATTPKQKPTKAPKKPTSTKKPKIMP 998
         |:| :||  ||
Db  1919 ----SPAPSTNTSAP-------CPETGP 1935
```

```
RESULT  15
ID   Q9XDH2    PRELIMINARY;    PRT;    763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH MUCIN HOMOLOG.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang Y., Moreno C., Singh M.;
RT   "Cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
```

```
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF071081, AAD41594.1; -.
DR   InterPro; IPR0029951; Atrophin.
DR   InterPro; IPR0029965; P_rich_extensn.
DR   InterPro; IPR003882; Pistil_extensin.
DR   PRINTS; PR01222; ATROPHIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE  763 AA;  75034 MW;  39168EC45A591GF8 CRC64;

STRAIN=H37RV;
```

```
Query Match             9.8%;    Score 715;   DB 2;   Length 763;
Best Local Similarity   28.2%;   Pred. No. 1.5e-39;
Matches  248;  Conservative  39;  Mismatches  361;  Indels  232;  Gaps  39;

Qy   288 PTPKAETTTKGPALTTPKEPTPTTTPKEPASTTPKEPTPTTTKSAPTTTPKEPAPTTTKSAP 347
         | |:   ||
Db     3 PVP---------------APRALAPLPPAPPAPAEPKSKPPFPPAP-----------PAPPCWMLVSAAP 46

Qy   348 TTPKEPAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTPKEPAPTTTKKPAPTTPKE 407
         |  | ||||
Db    47 PCP--PAP----PAPPKPKSKAPFPPVPPAPPARELAPLP--PAP------PEAPRE 90

Qy   408 PAPTTPKEPTTPTT----PKEPAPTTPKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK 461
         |:|  ||    ::|    ||||    :||  ||  |  |||        |||     | |
Db    91 SRPALPPCPPPEVVIPDPPEPAAPVPPRAENSPFPFPPPPARPKFVPPAPPVP--PVPNSPP 148

Qy   462 EP---APTTTKEPSPTTTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPSPTTTKEPAP 519
         :|    | |   |     |  | |||||  ||||||||||   ||||||
Db   149 FPPPPPAALNPPAP-----PAPPLANSPPLPPAPPPAGT--PPAAPWPPVPAAPKSKPA 201

Qy   520 TTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAP-----TTTKKPAPTAPKEEPAPTTPKETAP 575
         :|:   ||   ||||:|   |  | |:|         |||  |||  |:  ||:|||
Db   202 SPPRPPAP-----PMPATPMEFPPLPPVPPDPISKETPPAPPARPIPPAAPVPIPPVPPLP 256

Qy   576 TTPKKLTPTTPEKLAP---------TTPEKPAPTTPEELAPTTPAPPTTPEEPAPTT-- 625
         | |  | |  |||:||         :|| ||||   :|| | | |  |:  |:|
Db   257 PVPNKIPPAPP---APPVAVAAVLVAPCPLPPLPNNHPPAPPAAPVPGVPLAPLPNSHP 313

Qy   626 ---PKAAAPNTPKEPAP-------------------------ETTSKAPEN- 639
            ||||| |:|
Db   314 PAPPSAAPVGVPLAPLPISGRPVSVWKGSFTTLSTFCCRVCSGEVLAGALNPSRPSRSPL 373

Qy   640 --TTPKEPAPTTPKEPAPTTPKETA----PTTPKGTA--PTTLKEPAPTTP--KKPAPKE 689
           ||   | ||||||   |  |:|      |||||||   ||||||||||    |||||
Db   374 TTTTPALPAPIPPLPPLPPLPINTAVPPIPPLPPVTALAPLPPLAPLPISPGVPPAP-- 431

Qy   690 LAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPA------PTTPKEPAPTTPKGTAPTTL 743
         ||| ||| |||| :|||||||||||||||||||       ||||||||||||||||||||
Db   432 --PIPPGKPWTTPLPAPPAPEPK-TVPVLPPGPSCPPSERKNPPAPPEEPPEPKSSPALPP 488

Qy   744 KEPAPTTP----KKPAPKELAPTTPKGP-TSTTSDKPAPTTPKET--APTTPKEPAPTTP 796
         | | |||     ||||| ||||||     ||||||||||||||    |||||||||||
Db   489 APPAPSMPSAVRVPPSPPIPPAPPAAPRASMPALPPAPPSPPATRLCPPLPPSPPAPNSP 548

Qy   797 KKPAPTTPETPPPTTSEVSTPTTTKEPTTTIHKSP-----DESTPELSAEPTPKA-LENSP 850
         ||| |||| |||||:|    | |||   |||:|||      |||||| | |   |
Db   549 --PAPPAPTTPPKLLS--ANPCPVPPAPNRPAPPAPPAPPAPPELPAPPDPPTPVANSP 604

Qy   851 KEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTPAAPKMTKETATTTEKTTESKI 910
         ||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   605 --PAPPAPPAPPSALPFVNPPA-----------------------PPTPAAPK-------SRPAL 637

Qy   911 TATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATN 970
         ||||||||||||||
Db   638 PAAPPAPPAPPVRATTP--------------------------------PPAPPAPPAPN 665

Qy   971 SKATTPKPQKP----TKAPKKPTSTKKPKIMPKRVKRPKTTPTPTPRKMTSTM-----PELNP 1021
         | |||||| ||    |:|||
Db   666 SMALPPAPPDPPIPLLATPPAPAPPLPMSPPAAPLPPAAPDPAPPLTINQPPSPPLAP 725

Qy  1022 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET 1061
         ||  :|       ||       |:
Db   666 ...
```

Db   726 VPGAPLAPLPINGRPVFARKNSLI---------GSSSGDT 756

Search completed: April 26, 2002, 16:30:16
Job time: 631 sec

```
                    GenCore version 4.5
            Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:         April 26, 2002, 16:19:45 ; Search time 210.7 Seconds
                                            (without alignments)
                                            853.197 Million cell updates/sec

Title:          AA1
Perfect score:  6568
Sequence:       1 MAWKTLPIYLLLLSVFVIQ..........ARAITTRSGQTLSKVWYNCP 1229

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5
Searched:       473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters:          473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SPTREMBL_17:*
                1: sp_archea:*
                2: sp_bacteria:*
                3: sp_fungi:*
                4: sp_human:*
                5: sp_invertebrate:*
                6: sp_mammal:*
                7: sp_mhc:*
                8: sp_organelle:*
                9: sp_phage:*
                10: sp_plant:*
                11: sp_rodent:*
                12: sp_virus:*
                13: sp_vertebrate:*
                14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES

Result                  Query
No.     Score   Match Length DB ID         Description
------  -----   ------------ -- --          -----------
1       6470.5    98.5   1404  4  Q92954     Q92954  homo sapien
2       6455.5    98.3   1404  4  Q9N435     Q9N435  homo sapien
3       3173.5    48.1   1054 11  Q9BX49     Q9BX49  homo sapien
4       1713.5    26.1    401 10  Q9MM99     Q9MM99  mus musculu
5        969       14.8   1079  5  O77765     O77765  bos taurus
6        939.5     14.3   1049  5  Q9N4S7     Q9N4S7  caenorhabdi
7        865.5     13.2   1795  5  Q9IYS1     Q9IYS1  drosophila
8        853       13.0   1489 10  O76894     O76894  drosophila
9        808.5     12.3   1315 10  O96449     O96449  phytophthor
10       802       12.2   1274  5  Q9SPM0     Q9SPM0  zea mays (m
11       800.5     12.2   1480 10  Q20007     Q20007  caenorhabdi
12       800.5     12.0   1188 10  Q9LIE8     Q9LIE8  arabidopsis
13       784       11.9   2187 11  Q41805     Q41805  zea mays (m
14       778.5     11.9   2112  5  P70670     P70670  mus musculu
15       739       11.3    763  2  Q9VEL9     Q9VEL9  drosophila
16       715       10.9    555 10  Q9XDH2     Q9XDH2  mycobacteri
17       707.5     10.8   1151 13  Q5PQ6      Q5PQ6   chlamydomon
18       671       10.2   4880 11  O57580     O57580  gallus gall
19       667.5     10.2   5085 11  Q9IIT1     Q9IIT1  rattus norv
                  10.2          11  Q9JKS6     Q9JKS6  rattus norv
```

```
RESULT    1
ID     Q92954                PRELIMINARY;          PRT;    1404 AA.
AC     Q92954;
DT     01-FEB-1997 (TrEMBLrel. 02, Created)
DT     01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT     01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE     MEGAKARYOCYTE STIMULATING FACTOR.
OS     Homo sapiens (Human).
OC     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX     NCBI_TaxID=9606;
RN     [1]
RP     SEQUENCE FROM N.A.
RA     Turner K.J., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA     Merberg D.M., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA     FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA     Jacobs K., Turner K.;
RL     (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL     Mosher D.F. (eds.);
RL     Biology of vitronectins and their receptors., pp.45-52,
RL     Elsevier Science Publishers B.V. (1993).
RN     [2]
RP     SEQUENCE FROM N.A.
RA     Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA     Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA     Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA     Bhatia S., Kriz R., Hewick R., Clark S.C.;
RN     [3]
RP     SEQUENCE FROM N.A.
RA     Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA     Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA     Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RL     Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL     Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
EMBL; U70136; AAB09089.1; -.
DR     InterPro; IPR000585; Hemopexin.
DR     InterPro; IPR001212; Somatomedin_B.
DR     InterPro; IPR024100; GF_cysknot.
DR     Pfam; PF00045; hemopexin; 2.
DR     Pfam; PF01033; Somatomedin_B; 2.
DR     PRINTS; PR00438; GFCYSKNOT.
```

```
                                ALIGNMENTS

20   663.5   10.1   2284   5   Q9VPG1   Q9vpg1 drosophila
21   660.5   10.1   6677   5   Q9N435   Q9n435 caenorhabdi
22   659.5   10.0   1229   5   Q94185   Q94185 caenorhabdi
23   653.5    9.9   1514   5   Q9GUW7   Q9guw7 leishmania
24   647       9.9   4833  11   Q9QYX6   Q9qyx6 mus musculu
25   647       9.9   5038  11   Q9QYX7   Q9qyx7 mus musculu
26   643       9.8   2089   4   Q14676   Q14676 homo sapien
27   635       9.7   7962  10   Q10465   Q10465 homo sapien
28   633       9.6    489  10   Q41707   Q41707 vigna ungui
29   632       9.6    761  10   Q92QI0   Q92qi0 arabidopsis
30   632       9.6   6632   5   O17362   O17362 caenorhabdi
31   629       9.6   3507   5   Q23587   Q23587 caenorhabdi
32   628.5     9.6   1612   5   Q9VYQ2   Q9vyq2 drosophila
33   627       9.6    990  13   Q91803   Q91803 xenopus lae
34   625       9.5   6642   5   O01761   O01761 caenorhabdi
35   624.5     9.5   2768   5   Q9VC00   Q9vc00 drosophila
36   622.5     9.5    839   2   Q9RX57   Q9rx57 deinococcus
37   622       9.5   3570   4   Q99552   Q99552 homo sapien
38   620.5     9.4    971   5   Q9XVS4   Q9xvs4 caenorhabdi
39   615       9.4   2344   5   Q9N3Y8   Q9n3y8 caenorhabdi
40   607.5     9.2    801   5   O23635   O23635 caenorhabdi
41   607.5     9.2    924  12   Q99307   Q99307 epstein-bar
42   605       9.2    379   5   O27929   O27929 drosophila
43   600.5     9.1   1893   5   Q9NKC9   Q9nkc9 drosophila
44   598.5     9.1    409  10   Q9SBM1   Q9sbm1 volvox cart
45   577       8.8    956  10   Q9LJ64   Q9lj64 arabidopsis
```

```
DR   PRINTS; PR000222; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
SQ   SEQUENCE   1404 AA;  AABD7AD19B35F4F6 CRC64;

Query Match
Best Local Similarity   98.5%;   Score 6470.5;   DB 4;   Length 1404;
Matches 1229;   Conservative  87.5%;   Pred. No. 0;   Mismatches  0;   Indels 175;   Gaps  1;

Qy    1 MAWKTLPIYLILLLLSVFVIQQVSS----------------------------------- 24
         |||||||||||||||||||||||
Db    1 MAWKTLPIYLILLLLSVFVIQQVSSQDLSSCAGRGCEGYSRDATCNCDYNCQHYMECCPDF 60

Qy   25 ---------------------------------------------------------- 24

Db   61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 120

Qy   25 ---------------------------------------------------------- 24

Db  121 PPSGASQTIKSTTKRSPKPPNKKKKTKVIESEEITEEHSVSENQOESSSSSSSSSSSSTIW 180

Qy   25 --------------VKDNKKNRTKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST 65

Db  181 KIKSSKNSAANRELQKKLKVKDNKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST 240

Qy   66 TQHNKVSTSPKITPAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG 125
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TQHNKVSTSPKITPAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG 300

Qy  126 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTFGPALTTPKEPTPTTPKEPAS 185
         |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  301 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS 360

Qy  186 TTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTKSAPTTPKEPAPTTPKEPAPTTTKEP 245
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTKSAPTTPKEPAPTTPKEPAPTTTKEP 420

Qy  246 APTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKEPAPTTPKEPAPTTPK 305
         ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  421 APTTTKSAPTTPKEPAPTTPKEPAPTTPKRPAPTTPKEPAPTTTKEPAPTTPKEPAPTTPK 480

Qy  306 EPAPTTAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTTKEPAPT 365
         ||||||| |||||||||||||||||||||||||||||||||
Db  481 EPAPTTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTKSAPTTTKEPAPT 540

Qy  366 TTKSAPTTPKEPSPTTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKKP 425
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db  541 TTKSAPTTPKEPSPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKKP 600

Qy  426 APTAPKEPAPTTPKETAPTTPKKLIPTTPEKLAPTTPEELAPTTPEEPTPTT 485
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 APTAPKEPAPTTPKETAPTTPKKLIPTTPEKLAPTTPEELAPTTPEEPTPTT 660

Qy  486 PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 545
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 720

Qy  546 APTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKG 605
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 APTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKG 780

Qy  606 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPK 665
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPK 840

Qy  666 KPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT 725
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 KPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT 900

Qy  726 TKTPAATKPEMTTPAAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTQV 785
```

```
RESULT 2
Q9BX49
ID   Q9BX49          PRELIMINARY;    PRT;  1404 AA.
AC   Q9BX49;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).
GN   BG174L6.2
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133553; CAC36090.1; -.
SQ   SEQUENCE   1404 AA;  782A11746B3FDEE5 CRC64;

Query Match
Best Local Similarity   98.3%;   Score 6455.5;   DB 4;   Length 1404;
Matches 1226;   Conservative  87.3%;   Pred. No. 0;   Mismatches  0;   Indels 175;   Gaps  1;

Db  901 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTQV 960
         |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||

Qy  786 TSTTTQDTTPKITTLKTTTLAPKVTTTTKTTTTTEIMNKPEETAKPKDRATNSKATTTPK 845
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 TSTTTQDTTPKITTLKTTTLAPKVTTTTKTTTTTEIMNKPEETAKPKDRATNSKATTTPK 1020

Db  846 PQKPTKAPKKPPTSTKKPKTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 905
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1021 PQKPTKAPKKPPTSTKKPKTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080

Db  906 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 965
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1081 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1140

Db  966 DETNICNGKPVDGLTTTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFT 1025
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1141 DETNICNGKPVDGLTTTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFT 1200

Db 1026 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKFIFKGCGGLITGQIVAALSTAKYKMNPESVY 1085
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1201 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKFIFKGCGGLITGQIVAALSTAKYKMNPESVY 1260

Db 1086 FFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHIIRIQY 1145
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1261 FFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHIIRIQY 1320

Db 1146 SPARLAYQODKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV 1205
         ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Qy 1321 SPARLAYQODKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGIDYYAFSKDQYYNIDV 1380

Db 1206 PSRTARAITTRSGQTLLSKVWVNCP 1229
         |||||||||||||||||||||||||
Qy 1381 PSRTARAITTRSGQTLLSKVWVNCP 1404

Qy   25 ---------------------------------------------------------- 24

Db   61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 120

Qy   25 ---------------------------------------------------------- 24

Db  121 PPSGASQTIKSTTKRSPKPPNKKKKTKVIESEEITEEHSVSENQOESSSSSSSSSSSSTIR 180
```

```
Qy    1 MAWKTLPIYLILLLLSVFVIQQVSS----------------------------------- 24
         |||||||||||||||||||||||
Db    1 MAWKTLPIYLILLLLSVFVIQQVSSQDLSSCAGRGCEGYSRDATCNCDYNCQHYMECCPDF 60
```

```
Qy     25  ------------------------VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST    65
                                   ||||||||||||||||||||||||||||||||||||||||
Db    181  KIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST   240

Qy     66  TQHNKVSTSPKITTAKPINPRSLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDG   125
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TQHNKVSTSPKITTAKPINPRSLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDG   300

Qy    126  KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTKPEPTTTPKEPAS   185
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTKPEPTTTPKEPAS   360

Qy    186  TTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEP   245
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  TTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEP   420

Qy    246  APTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTTPK   305
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  APTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTTPK   480

Qy    306  EPAPTAPKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT   365
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  EPAPTAPKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT   540

Qy    366  TTKSAPTTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKKP   425
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  TTKSAPTTPKEPSPTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKKP   600

Qy    426  APTTAPKKEPAPTTPKETAPTTTPKKLTPTTPEKLAPTTTPEELAPTTTPEEPTTTT   485
                     |||||||||||||||||||||||||||||||||||||||||||||
Db    601  APTTPKEPAPTTPKETAPTTTPKKLTPTTPEKLAPTTTPEELAPTTTPEEPTTTT   660

Qy    486  PEEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKEPAPTTTPK   545
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  PEEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTTPKG   720

Qy    546  APTTPKKPAPKELAPTTTPKEPSTTSDKPAPTTPKGTAPTTTPKGTAPTTTPKG   605
                     |||||||||||||||||||||||||||||||||||||||||||||||
Db    721  APTTPKKPAPKELAPTTTKEPSTTCDKPAPTTPKGTAPTTTPKGTAPTTTPKG   780

Qy    606  TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTTPKGTAPTTTPK   665
                     |||||||||||||||||||||||||||||||||||||||||||||
Db    781  TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTTPKEPAPTTPK   840

Qy    666  KPAPTTPEETPPTTPEVSTPTTTPKEPTTIHHKSPDESTPELSAEPTPKALENSPKEPGVPT   725
                     |||||||||||||||||||||||||||||||||||||||||||||
Db    841  KPAPTTPETPPPTTSEVSTPTTTKEPTTIHHKSPDESTPELSAEPTPKALENSPKEPGVPT   900

Qy    726  TKTPAAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV   785
                     |||||||||||||||||||||||||||||||||||||||||||||||
Db    901  TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV   960

Qy    786  TSTTTQDTTPFKITTLKTTTLAPKVTTTKKTTTTELMNKPETAKPKDRAINSKAATTPK   845
                     |||||||||||||||||||||||||||||||||||||||||||||
Db    961  TSTTTQDTTPFKITTLKTTTLAPKVTTTKKTTTTELMNKPETAKPKDRAINSKAATTPK   1020

Qy    846  PQKPTKAPKKPTSTKPKTMPRVRKRKTTTPRKMTSTMPELNPTSRIAEAMLQTTTRPN   905
                     |||||||||||||||||||||||||||||||||||||||||
Db    1021 PQKPTKAPKKPTSTKKPKTMPRVRKPKTTTPRKMTSTMPELNPTSRIAEAMLQTTTRPN   1080

Qy    906  QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS   965
                     ||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS   1140

Qy    966  DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDVYFT   1025
                     |||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDVYFT   1200

Qy    1026 RCNCEGKTFFRKDSQVWRFTNDIKDAGYPKPIFKGFCGGLTGQIVAALSTAKYKNWPESVY   1085
                     ||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 RCNCEGKTFFFKDSQVWRFTNDIKDACYPKPIFKGFCGGLTGQIVAALSTAKYKNWPESVY   1260

Qy    1086 FFKRGGSIQQVIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY   1145
                     |||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 FFKRGGSIQQVIYKQEPVQKCPGRRPALNYPVYGETTQVRRRFERAIGPSQTHTIRIQY   1320

Qy    1146 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV   1205
                     |||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV   1380

Qy    1206 PSRTARAITTRSGQTLSKWWYNCP 1229
                     ||||||||||||||||||||||||
Db    1381 PSRTARAITTRSGQTLSKWWYNCP 1404
```

RESULT   3

Q9JM99

```
ID   Q9JM99        PRELIMINARY;      PRT;   1054 AA.
AC   Q9JM99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE   PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN   PRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ikegawa S., Nakamura Y.;
RT   "a novel mouse gene highly homologous to a human gene encoding
RT   megakaryocyte stimulating factor precursor and cartilage superficial
RT   zone protein.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034730; BAA92310.1; -.
DR   MGD; MGI:1891344; Prg4.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001212; Somatomedin_B.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00200; SOMATOMEDIN_B; 2.
DR   PROSITE; PS00524; P_RICH_EXTENSN; UNKNOWN_1.
SQ   SEQUENCE   1054 AA;  115991 MW;  4FC64BFA42B3235 CRC64;
```

```
Query Match             48.3%;  Score 3173.5;  DB 11;  Length 1054;
Best Local Similarity   53.3%;    Pred. No. 1.2e-204;
Matches   652; Conservative  52; Mismatches 150; Indels 369; Gaps 21;

Qy     19  IQVSVSVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVT--TPDTSTQHNKVSTSPK    76
               |:  |||||||  ||||||||||||||||||||||||||:|:||||||  |||||||
Db    189  LQKNPNVKDNKKNKTPKKKPNPEPPAVDEAGSGLDNGEFKLTPPPDPDPPTTPHSKVATSPK   248

Qy     77  ITTAKPINPRSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQ   136
               ||:||||||||||:|:||  ||:| ||:||  |:  ||:||:|:
Db    249  TTAAKPVTRKPSLAPNSETSKEASLASNKETTVATNKQSSA-SKKKTTSVKETR   307

Qy    137  SIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPASTTPKEPTPTTI   196
               |||||||||||||||||||||||:||   |:|:||||||||  ||||
Db    308  SAEKTSDKDV-----------------------------EPTSTTPK------   325

Qy    197  KSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTT   256
               |||||||||||  |:||||||||||||:|||  |||
Db    326  NSAPTTKPV--TTTKESKFLP----------------LPQEPEPTTPAKEPEPTTTKPEPPTT   370

Qy    257  PKEPAPTTPKKPAPTTPKEPAPTTTPKEPTTTPKKEPAPTTTPKEPAPTTPAPKKPA   316
               ||||  |:||||||||||||||:||||  ||||  ||||||||||||  ||||
Db    371  RKEEPPTTPKEPEPTTTPKEPEPTTPKEPEPTTTPKEPGPTTPKEPE   430

Qy    317  PTTPKEPAPTTPKKPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKE   376
```

```
Db    431  |||  |||  ||  ||||||||||||  |||||                         ||||||
           PTTKEPEPTTTKEPEPSTTPKKEPEPTTPKEP----------------EPTTPKE  468

Qy    377  PSPTTTKEPAPTTPKEPAPTTPKEPAPTTPKKPAPTTPKKPAPTAPKEPAPT        436
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    469  PEPTTLKEPEPTTPKEPEPTTPKEPEPTTPKEPEPTTPKEPEPTTPKEPEPT        512

Qy    437  TPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPTTPEEPAPTTPKA      496
           ||||||                                                ||||||
Db    513  ------------------------------------------EPTTPKE          512

Qy    497  AAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPK         556
           ||||||||||||||||||||||||||||||||||||||||
Db    513  ------------------------------KEPEPTTPKKKP---                549

Qy    557  ELAPTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKEPAP          616
           |||||||||||||||          |||||||||||||||||||
Db    550  ------------------EPTTPKEPVPTTPKEPEPTTP                      575

Qy    617  TPPKKPAPKELAPTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPPK        676
           ||||||          |||||||||          |||||||||||
Db    576  --TTPKEPEP--------------------TTPKEPEPTTPKEPEPTTPKKKEPEPTT    617

Qy    677  PTTSEVSPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEM   736
           ||||                                                 ||||
Db    618  TSPKT                                                        622

Qy    737  TTTPAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPF  796
           ||:|:|
Db    618  TSPKT                                                        622

Qy    797  KITTLKTTTLAPKVTTTKTTITTEIMNKPEETAKPKDRATNSKAT------TPKP-Q    847
           ||||||||||||           |||:                  |||:
Db    623  -TTLKATTLAPKVTAPAE------EIQNKPEETTPASEDSDDSKTTLKPQKPTKAPKPTK  675

Qy    848  KPTKAPKKPTSTKKPKTMPRVRKPKTTPPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT   907
           |||||||||||||||  ||||:||||| ||||:||| ||||||||||||
Db    676  KPTKAPKKPTSTKKPKT--PKTRKPKTTPAPLKTTSATPELNTTP--LEVMLPTTTIPKQT 732

Qy    908  PNSKLVEVNPKSEDAGGAGEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDE 967
           |:|| |||||||                                 |||||||||||||
Db    733  PNPETAEVNPDHEDADGGEGEKP-LIPGPPVLFPTAIPGTDLLAGRLNRGININPMPSDE  791

Qy    968  TNICNGKPVDGLITTLRNGTLVAFRGHYFWMLSPFSPSPSPARRITEVWGTPSPIDTVFTRC 1027
           |:|:||||||||||||||||||||||||||||:|||      |||||||||||||||||||
Db    792  TNICNGKPVDGLITTLRNGTLVAFRGHYFWMLNPFRPPSPPRRITEVWGTPSPIDTVFTRC  851

Qy   1028  NCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYFF  1087
           ||||||||||||||:|:|||::  ||||||::|| |||:|||||||||:|||:|||||||
Db    852  NCEGKTFFFKDSQYWRFINDVDPGYPKQIVKGEGGLTGKIVAALSIARYKDRPESVYFF    911

Qy   1088  KRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQYS-  1146
           ||| :|||||:|||:||| |||||:|||| ||||::||||||::||| :||||:|:||:
Db    912  KRGGNIQQYTYKQEPMKKCTGRRPAINYSVYGEAAQVRRRFERAVGPFQOTHIFRIHYSV  971

Qy   1147  PARLAYQDKGVLHNEVKYSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDVP  1206
           | ::|||||| ||||||:|::||| ||||||||:|||||||||:||||||||||:|||||
Db    972  PMRVSYQDKGFLHNEVKVSTMWRGFPNVVTSAITLPNIRKPDGTDYYAFSKDQYNIDVP   1031

Qy   1207  SRTARAITTRSGQTLSKVWYNCP                                       1229
           |:||||||||||||||:|||||:
Db   1032  TRTARAITTRSGQTLSKIWYNCP                                       1054
```

RESULT 4
ID   077765
AC   077765;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS   Bos taurus (Bovine).

```
                      PRELIMINARY;   PRT;   401 AA.
```

OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ARTICULAR CARTILAGE;
RA   Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA   Aydelotte M.B.;
RT   "Immunodetection and Partial cDNA Sequence of the Proteoglycan,
RT   Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT   Joints.";
RL   J. Orthop. Res. 0:0-0(1998).
DR   EMBL; AF056218; AAD13404.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   Pfam; PF00045; hemopexin; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   SMART; SM00120; HX; 2.
FT   NON_TER   1   1
FT   NON_TER   1   1
SQ   SEQUENCE   401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;

```
Query Match            26.1%;    Score 1713.5;  DB 6;   Length 401;
Best Local Similarity  80.2%;    Pred. No. 2.8e-107;
Matches  320;   Conservative  24;   Mismatches  52;   Indels  3;   Gaps  3;
```

```
Qy    832  PKDRATNSKAATTPKPQKPTKAPKKPTSTKKPKTMPRVRKPKTTPPRK-MTSTMPELNPT  890
           || |||||| |||||||||||||||||||||||| ||||||||||| | :||:||||||
Db      5  PKGRATNSQVTTPKPQKPTKAPKKPTSTKKPRT-PRVRKPKTTPTPPKTTTSAMPEPTPT   63

Qy    891  SRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGEGETPHMLLRPHVFMPEVTPDMDYL  950
           | :||||||||||| | ||| ::||:|| ||| |||| ||||::||:: |||:||
Db     64  S-LPEAMLQTTTRPPTPNSEIIDVNSENEDGDAAEGEKPHMIFRPPVLTPIVIPGTEII   122

Qy    951  PRVPNQGIIINPMLSDETNICNGKPVDGLITLRNGTLVAFRGHYFWMLSPFSPSPSPARRI 1010
           || |:|| ||||:||||||:|||:||||||||||||||||||||||||:||: :|| |||
Db    123  VRGPSQGFGINPMFSDETNLCNGRPVDGLITLRNGTLVAFRGHYFWMLTPFTPPPPRRI   182

Qy   1011  TEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVA  1070
           ||||| |||||||||||||||||||||
Db    183  TEVWGTPSPIDTVFTRCNCEGKTFFFKGSGSVQQYTYKQEPTQKCTGRRPAINYSVYGE    242

Qy   1071  ALSTAKYKNWPESVYFFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFE   1130
           ||| |:||: |||||||||| :|||| :|||:||| ||||||:||| ||| :||||||
Db    243  ALSIAQYKSRPESVYFFKRGGSVQQYTYKQEPTQKCTGRRPAINYSVYGETAQYVRRRFE  302

Qy   1131  RAIGPSQTHTIRIQYSPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGY  1190
           ||||||| | ||:|| || |||||| ||||||||:||||||||||||||||||||||||
Db    303  RAIGPSQVHIRIHYTPVRVPYQDKGFLHNEVKVSTLWRGLPNVVTSAISLPNIRKPDGY   362

Qy   1191  DYYAFSKDQYYNIDVPSRTARAITTRSGQTLSKVWYNCP                       1229
           |||| |||||||||||||||||||||||||||: |||||
Db    363  DYYALSKDQYYNIDVPSRTARAITTRSGQTLSNTWYNCP                       401
```

RESULT 5
ID   Q9N4S7
AC   Q9N4S7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y51B11A.1 PROTEIN.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;

```
                      PRELIMINARY;   PRT;   1079 AA.
```

```
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Cotton M.;
RT   "The sequence of C. elegans cosmid Y51B11A.";
RN   [3]
RP   SEQUENCE FROM N.A.
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   1079 AA;   110532 MW;   8DBDE38824CF80CA1 CRC64;

Query Match             14.8%;  Score 969;  DB 5;  Length 1079;
Best Local Similarity   29.9%;  Pred. No. 6.3e-57;
Matches  327;  Conservative  137;  Mismatches  472;  Indels  156;  Gaps  44;

Qy     22 VSSVKDNKNRTKKPTPKRPPVVDEAGSGLDNGDFKVTT---PDTSTTQHNKYSTSP-KI 77
             ||:   ::   |  |||        :  :::        :|:|.:|  |:::   :|
Db     62 LSSTPTSSSTPIKETTTAPETTISTEPPSSSTTPVQTTT-TTAPETTISTEAPSSSTTPVQT 121

Qy     78 TTAKPINPRPSLPPNSDTS--KETSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKET 135
             || :           | :: |  |  |:||:   |:|||   |:|  | |||:|  |
Db    122 TTTTAPETTISTEPPSSSTTPVQTTTTTAPETTISTEPPSSSTTPVQTTT----TTTAPET 175

Qy    136 QSIEKTSAKDLAPTSKVLAKPTPKAETTTKGP - ALTTPKEPPTTPKEPASTTPKEPTP 193
             :|:   |   ||| |  || :|  |:||   |   ::||||| ||||||| ||||| |
Db    176 TSTEPPSSS----TSPVQTTTTTAPETTISTEPPSSSTTPVQTTTTTAPETTISTEPPSSST 231

Qy    194 TTIKSAPTTPKEPAPTTTPKKSAPTTPKEPAPTTTPKEPAPTTTKEPAPTTTKSA 253
             |  ||| |||||| |||  |   | |||||| ||   ||||  |  ||| || |
Db    232 TPVQTTTTTAPETTISTEPPSSSTTPVQTTTTTAPE---TTISTEP-PSSSTTPVQTTTTA 287

Qy    254 PTTPKEPAPTTPKKPAPTTPKEPAPTTTKEPAPTTTKEPAP--TTPKEPAPTAP 312
             | ||||||| ||   ||| |||||| |||:||  | :|||  || ||||| |
Db    288 PETTISTEPPSSSTTPVQTTTTTAPETTISTEP-PSSSTTPVQTTTTTAPETTISTEPPSSS- 345

Qy    313 KKPAPTTPKEPAPTTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAP-TTTKEPAPTTKSA 370
             ||| ||||||| |||   ||||  |  |  ||||||| ||| |   ||||||| || :
Db    346 -----TTPVQTTTTT----APETTRTEPPSSSTTPVQNTTTTAPETTISTEPPSSSTTPVQ 396

Qy    371 PTTPKEPSPTTPKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKKPAPTAP 430
             | ||||| |||||| ||| |||||| ||   ||| |||||| ||  ||| |
Db    397 TTTTTAPETTISTEPPSSST---TPVQTTTTTAPETTISTEP-PSSSTTPVQTTTTTAPETTS 453

Qy    431 KEP---APTTPKETAPTTPKKLIPTTPEKLAPTTPKEPAPTTTPPEELAPTTPKEP--TPTTP 486
             |:|   || |||| ||| |  : |||| ||||||||| |||   |  ||||||||  | ||
Db    454 TEPPSSSTTPVQTTTTTAPETTISTEP---PSSSTTPVQTTTTTAPETTISTEPPSSSTTP 509

Qy    487 EEPAPTTPKAAAPNTPKEPAP---TTPKEPAPTTPKEPAPTTPKGTAPTTLK 543
             : ||| ||  |||  ||| |   | ||||||| ||||||| || |||| ||
Db    510 VQTTTTT----APETTISTEPSSSSTTPVQTTTTTAPETTISTEPPSSSTTPVQTTTTTTAP 564

Qy    544 EPAPTTPKKPAPKELAPTTKEPTSTTSDKP-APTTPKGTAPTTPKEPAPTTPKEP--AP 600
             | ||| || || : ||| |||  |||    || |||| ||| |||||| ||||||  ||
Db    565 ETTTSTEPPSSSTTPVQTTTTTAPETTISTEPPSSSTTPVQTTTTTAPE---TTSTEPPSSS 621

Qy    601 TTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPISTTSDKP-APTTPKETAPTTPKEP 659
             | || |||| ||||||| ||   || |||| |  |  |||    || |||| |||||||
Db    622 TTPVQTTTTTAPETTISTEPPSSSTTPVQTTTTTTAPETTISTEPPSSSTTPVQTTTTTAPET 681

Qy    660 APTTPKKPAPTTPETPPPTTSEVSTPTTPKEPTTIHKSPDSTSPELSAEPTPKALENSPK 719
             | || ||||| |||  | | :: | | |||| ||:|| |  |:||  | | |||  |||
Db    682 TSTEP---PSSSNTPVQTTTTAPETTISTEPPSSSTTPVQTTTTTAPETTI-----STE 731

Qy    720 EPGVPTT--KTPAATKPEMTTPAKDKTTERDLRTT-----PETTTAAPKMTKETATTTEKT 773
             |   || |||   |  :  ||  | ||:| |  |     ||| | |||||| |||| |||
Db    732 PPSSSTTPVQTTTTTIIAPETTISTEPPSSSTTPVQTTTTTTAPETTISTEPPSSSTTPVQTTTT 791

Qy    774 TESKITAPTTQVTSTTTQDTTTPFKIITTLKTTTLAPKVTTTTKKTITTTEIMNKPEETA-KP 832
             | :||||||| |||| ||  ||   | | |||| || ||||  | ||||||| | |||| ||
Db    792 TAPETTISTEPPSSSTTPVQTTT--TTTAPETTISTEPPSSSNTPVQTTT--TTAPETTSTEP 847

Qy    833 KDRATNSKATTPKPQKPTKAPKKPTSTTKPKPKTMPRVRVKTKTTTPPTPRKMTSTMPELNPTSR 892
             |  | :|| ||| | | | || ||||| || | ||  | ||||| |||| ||| | |||: |||
Db    848 PSSSTISPVQTTTTTTAPETTISTEPPSSSTTPVQTTTTTTAPETTISTEPPSSSTTPVQTTTTTT 907

Qy    893 IAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEG-ETPHMLRPHVFMPEVTP---DMD 948
             |||  |||| ||  || |   | || ||  | ||  ||| || | | ||| |   |||
Db    908 APET---TSTEP---PSSSTTPVQTTTTTAPETTISTEPPSSSSTTPVQTTTTTVPCDCSLS 961

Qy    949 YLPRV--------PNQGIIINPMLSDETN------------ICNGKPVDGLTT 981
             | |:         || |: ||| | :|               ||| | |||
Db    962 YIDRVVYPFTTEWWENKRDIIIQSYDSPRRTAFSFSVSTPDIGTCTAILCIYTSEGGISN 1021

Qy    982 LRNGTLVAFRGHYFWMLS-------PFSPSPARRITEVWGIPSPIDTVFTRCNCEGKTF 1034
             | |  |||  ||:|| |        || | | |  | ||  | | | | |||||| |
Db   1022 L-NATL-----FIGLSDGSSIDLPF-------YNVPTGLE-IWPEINCEGNK- 1059

Qy   1035 FFKDSQYWRTN 1046
             :  |  ||:||
Db   1060 ------WSYNN 1064
```

```
RESULT 6
ID   Q9I7S1       PRELIMINARY;       PRT;   1049 AA.
AC   Q9I7S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helman T.J., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
```

```
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003495; AAG22353.1; -.
DR   FlyBase; FBgn0030561; CG5228.
SQ   SEQUENCE   1049 AA;   107278 MW;   954D629E7619671 CRC64;

Query Match                      14.3%;   Score 939.5;   DB 5;   Length 1049;
Best Local Similarity            34.6%;   Pred. No. 5.8e-55;
Matches  373;   Conservative  46;   Mismatches  443;   Indels  217;   Gaps  54;

Qy       19 IQQVSSVKDNKKNRIKKKPPTPKPPVVDEAGSGLDNGDFKVTTPDTS--TTQHNKVSTSP   75
            |||    : |||         |||  | |  |||:  |||||   |::   ||:|  |  :
Db       23 IRIVPGQKPNEIGTTAKPTTLKP--------------------------TEGTSAKPTTLKPTEGTSA   64

Qy       76 KITTAKP-----INPRSLPPNSDTSKEFTSL------TVNKETTVETKETTTTNKQT--ST   123
            ||||||       |:  |   | | :  |        | |: | |:|||| |||    ||
Db       23 IRIVPGQKPNEIGTTAKPTTLKP--------------------------TEGTSAKPTTLKPTEGTSA   64

Qy       76 KITTAKP-----INPRSLPPNSDTSKEFTSL------TVNKETTVETKETTTTNKQT--ST   123
            : |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db       65 KPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPT   124

Qy      124 DGKEKTTSAKET--QSIEKTSAK--DIAPTSKVLAKPT-------PKAETTKGPALTTP   172
            :|        ||| |   | ||   | |  |   |        |||  |||  ||
Db      125 EG----TSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTS   180

Qy      173 KEPPPTTPKE---------PASTTPKEPT----PTTLKSAPTTPKEPAPTTTKSAPTTPKEPAPT   224
            :||     |||       ||||||||    |||||||||||||||||||||||||||
Db      181 AKPTTLKPTEGTSAKPTTLK--PTEGTTAKPTTLK--PTEGTSAKPT   236

Qy      225 TTTKEPAPTTPKEPAPTTTPKEPAPTTTKSAPTTPKE--PAPTTPKKPAPTTTPKEPAPTTTPK   282
            ||     |||||||||||||||||:|||||||| |     | |||||||||||||||||||||
Db      237 TLKPTEGTTAK---PTTLKPTEGTTAK-----PTTLNPTEGTTAKPTTLKPTEGTTAKPTTLNPT   293

Qy      283 EPT----PTTPKEPAPTTKEPAPTTPKEPAPTTAPKKPAPTTPKEPAPTTPK-----EPAPT   334
            |||    || ||||||||||||||||||||       ||||||||||||||     ||||
Db      294 EGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPT   353

Qy      335 TTTKEPSPTTPKEPAPTT-------TKSAPTTTKEPAPTTIKSA----PTTPKEPSPTTTKEP   385
            :|         |||||||       || |||||||||||| ||    |||||||||:|||
Db      354 TLKPTDGTTAK---PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPT   410

Qy      386 APTTTKEPAPTTPKKPAPTTPKEPAPTTTPKEPAPTTTTKKPA---PTAPKEPAPTTPKETA   442
            |||||||||||||||||||||||||||||||:||||||||||   ||  |||||||||| |
Db      411 EGTTAK---PTTLKPTEGTTAK---PTTLK-PTEGTSAKPTTLKPTEGTTAKPTTLKPTE   463

Qy      443 PTT---PKKLTPT-------TPEKLAPTTPKEPAPTTPEELAPT---TPEEPTPTTPEEPAPTT   493
            |||   |||  |||       |||: || |||||||| |||  |    |||| | ||||||||||
Db      464 GTTAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTAKPT   523

Qy      494 PKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTA--------PTTLK   543
            || |  ||||||||||||||||||||||||||||||||||||||        |||||
Db      524 PTTLKPTEGTTAKPTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLK   580

Qy      544 EPAPTTPK----KPA-------PKELAPT---------TTKEPTSTTSDKPA---------   575
            |  ||||||    |||       ||:|||         |||||||| ||||
Db      581 PTEGTTAKPTTLKPTEGTSAQTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTSAQ   640

Qy      576 PTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPK----KPAPKELAPTT   631
            ||||||||||||||||||||||   |||||||||||  |||||||    ||| || ||||
Db      581 PTEGTTAKPTTLKPTEGTSAQTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTSAQ   640

Qy      576 PTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPK----KPAPKELAPTT   631
            ||||| |||||||:||||||||:|||||||||||||| |||||||||||||||||||||
Db      641 PTTLKPTEGTTAK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAKPT   694

Qy      632 TKGPISTTSDKPAPTTPKE--TAPTTPKEPAPTTPKKPA---PTTPETPPTTSEVSTPTT   686
            : |         |  |||   ||||||||||||||||||   ||||||||||:||:|||
Db      641 PTTLKPTEGTTAK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAKPT   694

Qy      687 -TTKEPTTIHKSPDESTPELSAEPTP--------------------------KALENSPKEPGV--PTTKT-   728
                   :|||   :|    |   ||                         ||||| |||||  ||||
Db      695 TLNPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGT   754
```

RESULT 7
O76894

```
ID   O76894           PRELIMINARY;     PRT;   1795 AA.
AC   O76894;
DT   01-NOV-1998 (TrEMBLrel.  08,  Created)
DT   01-NOV-1999 (TrEMBLrel.  12,  Last sequence update)
DT   01-JUN-2001 (TrEMBLrel.  17,  Last annotation update)
DE   EG:5627.1 PROTEIN.
GN   EG:5627.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhou X., Zhu S., Zhu X., Smith H.O.,
```

```
Db      755 RTTAKPTTL--KPTEGT--SAKPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPTERTS   809

Qy      729 --PAATKPEMTTTAKDKTTERDLKTT--------PETTTAAPKMTKEFATT-------   769
            |  ||    | | | |         ::                 |  |  |
Db      755 RTTAKPTTL--KPTEGT--SAKPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPTERTS   809

Qy      810 AQPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKP   869
            ||||||||||||||||||||||||||||||||||: |||||: |||||||||||
Db      810 AQPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKP   869

Qy      770 TEKTTESKITATTTQVTSTTTQDTTPFKITTLKTTLAPKVTTTKK--TITTTEIMNKPE   827
            |: |||||||||: ||  |:  |  ||  |||||::|| ||||     |||||||||
Db      810 AQPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKP   869

Qy      870 TEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLKPTEGTSAKP   929
            |||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      870 TEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLKPTEGTSAKP   929

Qy      828 ETAKPKDRATNSKATTPKPQKPTKAPK----KPT--STKKPKTMPRVRKPKTTPT---PRK   879
            |: |    :  | : |  | |: |         |    ||:  |  |  | | |      |
Db      930 TTLKPTE-GTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTE   988

Qy      880 MTSTMP-ELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEG-ETPHMLLRP   936
             :  :    |||  :  | |  |  |   ||| |  | ||| |   |    ||  ||||
Db      930 TTLKPTE-GTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTE   988

Qy      989 RTSAQPTTLKPTERTSAQ--PTTLKPTEGTTAKPTTLKPTKGTSGQANNFETKRKERRP   1045
```

```
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.;
RT  "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL  Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RA  Benos P.;
RL  Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AE003421; AAF45644.1; -.
DR  EMBL; AL031028; CAA19845.2; -.
DR  FlyBase; FBgn0025390; EG:5G7.1.
DR  InterPro; IPR002557; Chitin_binding.
DR  InterPro; IPR002965; P_rich_extensn.
DR  Pfam; PF01607; Chitin_bind_2; 2.
DR  PRINTS; PR01217; PRICHEXTENSN.
DR  SMART; SM00494; ChtBD2; 2.
SQ  SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match     13.2%;  Score 865.5;  DB 5;  Length 1795;
Best Local Similarity  28.5%;  Pred. No. 9.3e-50;
Matches 276;  Conservative 129;  Mismatches 447;  Indels 115;  Gaps 28;

Qy   30 KNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTQHNKVSTSPKITTAKPINPRPSL   89
Db  431 KNDVTEAPEIKSPL----KLHLSENIVILPETTTT--TTTTIKPVVLTCPTISPPDTT   483

Qy   90 PPNSDT----------SKETSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSI  138
Db  484 PKPSTTTAVTKSTPKISSTEQHSTTTAKTITTKRPTTVIEKTSSATEKPRTTVVTTTQK  543

Qy  139 EKTSAKDLAPTSKVLAKPTPKAETTTKGPALTITKPKEPTTPTPKEPASTTPKEPTTIKS  198
Db  544 RSTTHNTSPDTKTTIRSTTLLSPKTTTPSTTTPSTTTPST----TTPSTTTPST---   594

Qy  199 APTTPKEPAPTTTKSAPTTPKEPAPTTTK-EPAPTTPKEPAPTTIKEPAPTTIKSAPTTP  257
Db  595 --TTPSTTTPSTTTVVSTHRPRTTSQKTTTASTTTKK----TTTSPKTTKTTDIPTST   648

Qy  258 KEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTAPKKPAP  317
Db  649 TSKLSTTTQKTTTTTTHKFTAATTSTEKPKTTTEKTSTVSTTTKKSTESSSPKPTSSTGKPT  708

Qy  318 TTPKEPAPTTPKEPAPTTTPKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEP  377
Db  709 TTPKPSTRTPTPTTPKVTTTTQITTTPLRSSTETISTQPPTTTTPQPTTTTTLTVTPKTS   768

Qy  378 SPTTTKEPAPTTPKEPAPTTPK----------KPAPTTPKEPAP------TTPKE   416
Db  769 TTTTTTTEKPTSSPKPTTTTQKTTSTAPNTKVAITTQKETTTPTQSTTSTTIFTRKTTTNN  828

Qy  417 PAPTTTKKPAPTAPKEPAPTTPK-ETAPTTPKKLIPTTPEKLAPTTPEKP------AP   467
Db  829 PEPTSTEKPITSTTTPKPSTTSVASSTEKTTISSPKPTTNSVKTSAL              888

Qy  468 TTPEELAPTTPEEPPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPT--TPKEPAPTT   525
Db  889 TSSTQRATSTTSSEPKTT-QNITTTTPKPTTLKTSTQEATTSTQKVSTVITTKKATESS   947

Qy  526 PKETAPTTPKGTAPTTLKEPAPTTPKPAPKELAPTTTKEPT--SSTSDKPAPTTPKGTA   583
Db  948 PLTTLSTEEPNTTPKPLRTTTPTTTSVTATTRITTTTISESSTETTTSTQKPKSTTPTTT 1007

Qy  584 PTTPKEPAPTTPKEGTAPTTLKEPAPTTPKPAPKELAP-TTTKGPISTTSDK         642
Db 1008 RTTPKVTVTVISTQNPTTTTSKTSTVTI----TTP-NPSSTQRPTTTTRQPTSTTAST  1061

Qy  643 PAPTTPKETAPTTPKEPAPTTPETTPPPTTSEVSTPTTT-TKE----------        690
```

```
Db 1062 TSIGTTRIPTTNPQNSTSSSTDLTTVTRPPCPDPDSTSDKNTNTACTQELQQVNLLELQS 1121
Qy  691 --------PTTIHKSPDESTPELSAEPTTKALENSPKEPGVPTTTTKTPAATKPEMTTTAKD  742
Db 1122 PQKQEQFTHTRTHTALTGSRNTLGGQEVPDYMDDAPSAAEAESGQATTPAKAPTTMSTLAAA 1181
Qy  743 KTTER---DLRTTPETTTPAAPKMTKETATTEKTTESKTITATTTQVTSTTTQDTTPEKIT  799
Db 1182 HLLQKLFHIISTTPPSREHAP--TQRPSSSOPSSSOSR-GVTTAQMARINLATSKPFIAH  1238

Qy  800 TLKTT-----TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKA-PK  854
Db 1239 SLRLSIQQLASTQKRSIPPKTLVTHNTTKEPED-SEYYDSETSEQYTDEDNEVLDKTQPR  1297

Qy  855 KPTSTKKPKTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVE  914
Db 1298 AMSSTTVAAVLPAVPSTTEREPQK-TSSSP---SPT----KATSSTTTQPIETTTGDLEY 1350

Qy  915 VNPKSED  921
Db 1351 DSSGSSD 1357

RESULT    8
96449
ID   96449;   PRELIMINARY;   PRT;   1489 AA.
AC   96449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RACE 1-11;
RA   Goernhardt B.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061185; AAC72308.1; -.
SQ   SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

Query Match     13.0%;  Score 853;  DB 10;  Length 1489;
Best Local Similarity  32.3%;  Pred. No. 5.3e-49;
Matches 355;  Conservative 46;  Mismatches 475;  Indels 222;  Gaps 49;

Qy   33 TKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTT-----QHNKVSTSPKITTAKPINPR   86
Db  338 TPYAPTEKPYDVEETTVVTEESTYAPTPKSETNAPTERMHYAHIEKPCDTEVTMYAPTEET  397

Qy   87 PSLP----------PNSDT----SKETSLTVNKETTVETKETTT---TNKQTSTD     124
Db  398 TYAPTEETTYAPTEETTYAPTEETTPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAP  457

Qy  125 GKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTKAETTTKGPALTTPKEPPTTPKEPA    184
Db  458 TEETTYAPTEETEPYEPTEETTYAPTEKTTYAPT---EETTYASTEETTYAPTEETTYAPA  514

Qy  185 STTPKEPTPTTIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT--TTKEPAPTTPKEPAPTT  241
Db  515 EETPYEPTEET-TYAPTEETTYAPTEETTYAPTEETTYAPAETTYAPAEETPYEPTEET    573

Qy  242 TKREPAPTTTKSAPT---------TPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPKE  291
Db  574 TYAPTEETT-YAPTEETTMYAPIEETTVYAPTEETTYAPAAEETPYEPTEETTYAPTEE   632

Qy  292 P-----------AP---TTKEPAPTTPKEP-----APTAPKKPAPTTPKEPARPTTP   328
Db  633 PTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEE        692
```

```
RESULT  9

ID   Q9SPM0      PRELIMINARY;    PRT;   1315 AA.
AC   Q9SPM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN.
GN   PEX2.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=POLLEN;
RA   Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RA   Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RT   "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159297; AAD55980.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR001998; Xylose_isom.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003592; LRR.out.
DR   Pfam; PF005560; LRR; 3.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR   SMART; SM00370; LRR; 3.
SQ   SEQUENCE   1315 AA;  134401 MW;  64C97A2A01F09336F CRC64;

Query Match             12.3%;  Score 808.5;  DB 10;  Length 1315;
Best Local Similarity   25.0%;  Pred. No. 4.4e-46;
Matches  255;  Conservative   88;  Mismatches  459;  Indels  219;  Gaps  33;
```

```
Qy   329 KEPAPT--TTKEPSPTTPKEP-------APT-TTKSAPT------------TTKEP------A 363
             |||            |||                |||             |||       |
Db   693 TTYAPTEETTYAPAEETYEPYEETTYAPTEETTYAPIEETTYAPIEETTYGPTEETTYA 752

Qy   364 PT-TTKSAPT--TPKEPSPTTTKEP------APTTPKEPAPTTPKKPAPTTPKEPAPT-- 412
             ||| |         |||            |||          |||
Db   753 PTEATTYAPTEETTYAPTEETTYEPIGETTYAPTEETTYAPTEETTYAPTEETTYAPTEE 812

Qy   413 TPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEE 472
             | | ||| |        |||      |||
Db   813 TPYEPTEETTYAPTEETTYEPTEET--TYTPTEETTYAPTEETTYAPEKTTYAPTEETT 870

Qy   473 LAPT--TPEEEPTPTTPEEP--------APT----------TPKAAAPNTPKEPAPTTPKEP- 513
           ||||                     |||                            |||
Db   871 YAPTEETTYEPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPT 930

Qy   514 ---APTTPKEPAPTTPKEAPTTPKGTAPT--TLKEPAPTTPKKKPAPK-ELAPT--TT 563
            |||             |||                 |||               ||
Db   931 EETTYAPTEETTYAPTEETTYAPAEETYAPAEETPYEPTEETTYAPTEETT 990

Qy   564 KEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP-----APTTPKGTAPT-------- 609
             |||             |||                      |||
Db   991 YAPTEETMYAPIEETTYAPIEETTYAPTEETTYAPTEETTYAST 1050

Qy   610 -------TLKEPAPTTPKKP-----APKE--LAPT--TTKGPTSTTSDKPAPTTP 648
                   |||               ||||  ||||
Db  1051 EETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETT 1110

Qy   649 KETAPTTPKEPAPTTPKKPAPTTPEPTPPTTSEVSTPTTTKEPT--TIHKSPDESTPELS 706
             |||          |||                         |||
Db  1111 YAPAEETPYEPTEETTYAPIEETMYAPIEETTYGPTEATTYAP 1169

Qy   707 AEPTPKA-LENSPKEPG------VPTTKTPAATKPEMTTTAKDKTTERDLRTTP----ETT 756
                                                                         |||
Db  1170 TEETTYAPTEETTYEPTCETTYAPTEETTYAPTEETTYEPTEET 1229

Qy   757 TAAPKMTKETATTTEKTTESKITATTTCVTSTTTQDTT----------PFKITT--LKT 803
                                                           |||
Db  1230 TYAP----TEETTYEPTEETTYAPTEETTYAPIDETTYGPTEE 1284

Qy   804 TTLAPKVTTTKKITTTTEIMNKPEETAKPKDRAT--NSKATTPKP-OKPTKAPKKPT--- 857
                                                        |||
Db  1285 TTYAPTEATTYAPTEETTYAPTEETTYEPTGETTYAPTEETTYAPMEETPYE 1344

Qy   858 --------STKKP-KIMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP--N 905
Db  1345 PAEESTSTVSTEKPCNTEEFTDEPTDEEPTDE--PSDEPTDEPTDEPTLLPTDEPSTPCDN 1402

Qy   906 QTPNSKLVEVNPKSEDAG 923
              :||
Db  1403 QGINGIGVENKVRYNNAG 1420
```

```
Qy    37 PTP-------------KPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPR 86
            |||
Db   412 PTPGGPPSSSVPGKPPSVPG-------KPAAPAPMPTPHTPPDVSP----EPL-PE 455

Qy    87 PSLPP----------NSDTSKETSLITVNKETTVETKETTTTNKQTTSDGKEKTTSAKETQ 136
            :|||
Db   456 PSPVPAPAPMRMPTLRSPPADEYIPT---PPVPAKSPPGTSPPASRGAPPLQOQPPAAS 511

Qy   137 SIEKTSAKDLAPTSKV--------------LAKPTPKAETTTTKGPALTTTPKEPTTTTKE 182
                                          :|:|
Db   512 SPPAPPVKSSPPPAAVVLPPPAKTPSPPAPVASPPEAPVSSPPQOVKKSPPPAPVASPP 571

Qy   183 PASTTPKEPTPTTIKSAPTTPKEPAPTTTPKSAPTTP-KEPAP------TTTKEPAPTTP 234
            |:|                         |:|
Db   572 PPMKSP--PPPARVASPPPLMKSPPPPAPVASPPQPLKSPPPPVLWLSTPSVKSPPPPVP 629

Qy   235 KEPAPTTTKEPAPTTTTKSAPTTPKE-----PAP----TTP----KKPAPTTPKEPAPTTPKE- 283
            |:|                      :|:      |||      |||
Db   630 VASPPPPVKSPPPLAPVSSPPSPPVKLPPLPAPGKSTPPEEEKPTPPEVKSSPPPEKSL 689

Qy   284 PTPTTPKEPAPTTKEPAPTTPKEPAPTTAKKPAPTTPKEPAPTTPKEPSPTT 343
            |||                              |:|
Db   690 PPPLLTTSPPPQEKPTPPSTPSKKPPPSPVETLPPPSKSSPEEPVSSPPQOAPKSSSP-- 747

Qy   344 PKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAP 403
            |:|                                |:|
Db   748 ---PAPVS--SPPPLKSSPPPPVPESSPPPPTPKSSPPLAPVSSPPQVEKTSPPPAPVSSPP 802

Qy   404 TTPKEPAPTTPKEPAPTTPKKPAPTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPE 463
              |:|                              |:|
Db   803 PTPKSSPPLAPVSSPPQVEKTSPPPAPVSSPPPAPVSSPPQVEKTSPPPAPV 862

Qy   464 KPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAP 523
            |||      :|:                             |||
Db   863 SSPPLEPKPSSPPPSVSSPPTTVKSSPPPTLSSPPMTRKSSSPPAHVSSPPEAEKSSPP 922

Qy   524 TTPKETAPTTPKGTA--------PTTLKEPAPTTPKKPAPKELAPTTPKEPTSTTSD 572
            ||:                     |:|
Db   923 LAPISSPPSEPKSPSSPPMVEKTSPPPAIVSSPPP---APVSSPPPVVKSSP 979

Qy   573 KPAPT------TPKGTAPTTP-----KEPAPTTPKEPAPTTPK--GTAPTTLKEPAPT 617
            ||||                      |||
Db   980 PPAPVSSPPPTTPKPLPPPAPVSSPPPPVVKSSPPPPVSSSPPPPVKSSPPPPVSSPPPT 1039

Qy   618 TPKKPAPKELAPTTTKGPISTTSDKPAPTT--PKETAPTTPK----EPAPTTPKKPAPTT 671
            |||                             |:|
Db  1040 XKPLPPP--APVSSPPPVVKSSPPPAPVSLPPPPTPRKPSPPRTRVSSSPPRVVKCCPPPL 1096

Qy   672 PETPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA 731
            |||                              |:|
Db  1097 VSSPPPAPKSLPPPTPVSSP--------PPEVKSSPPPTPVSSPPPAPKSSSPPPTPVS 1146

Qy   732 TKPEMTTTAKDKTTERDLRTTP------ETTTAAPKMTKETATTTEKTTESKITATTTQVT 786
            :|:|::|                     :|:::|
Db  1147 SPPP----------ELKSSPPPAPVSSPPSAPKSSPPPAPVSLPPPEVKSSPPPAPIS 1194
```

```
Qy    787  STTTQDTTPFKITTLKTTTLAPKVTTTKKTTITTEIMNKPEETAKPKIDRATNSKATTPKP  846
                :|                            :  |  ::|
Db   1195  SPPPPAKSP----------------------------PMSSLPPPVKSPPPPAPVSSP  1228

Qy    847  QKPTKAPKKPPTSTKKPKTMPRVRKPKTTPTPTRKMTSTMPELNPTSRIAEAMLQTTTRPNQ  906
                :|:|                           :|:|  :::  |:|
Db   1229  PPPMKSPPPPAPISSPPPAP--VKPPSLPPPAPVSSPPPAV-------TSAP--  1271

Qy    907  TPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD  966
                                      ||     |  |  ::|:
Db   1272  ----------------------PKKEEDSTA----------------------  1302

Qy    967  E 967
             :
Db   1303  K 1303
```

RESULT 10

Q20007

ID   Q20007;   PRELIMINARY;   PRT;   1274 AA.
AC   Q20007;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   COSMID F35A5.
GN   F35A5.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sulston J., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaughin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RL   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Leimbach D.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U46675; AAB52641.1; -.
SQ   SEQUENCE   1274 AA; 138065 MW; 07F6B0D29ZC1799F CRC64;

```
Qy     75  PKITTAKPI-NPRPSLPPPNSDTISKEISLTVNKETTVTETKETTTNKQTSTDGSKEKTTSAK  133
                :::                      :|  :::
Db    334  SPVVPPTPVKNPVKVKPPWEVDDEPAEEEVKPSAPEKKTTPVLKREPEPSSTTPSSSDPS  393

Qy    134  ETQSIEKTSAKDLAPTSKVLAKPTPKAE-----------TTTKGPA---------  172
                :|:|                 :|  :||
Db    394  PKKAAPAVKPRDSSPKKATPLQADPKAQAEVPPTPVKNPVKYKPPWEVDDEDPVEEVKQP  453

Qy    173  KEPTPTTP----KEPASTTPKEP-----TP-TTIKSAPTTPKEPAPTTTKSAPTTPKEPAP  223
                |:|:                ::|       :|:|
Db    454  EAPAKKTPVLKRKEPAAKDTAKPATSKTPETPEKKDPVKPRDSSPKKVAAKPDSAQAPA-  512

Qy    224  TTTKEPA--------------------  257
                ||    |:
Db    513  TPVKNPVKWRPPWEDDETPADDVSKPTDAKKTPSLAKKDPAPAKESLKPRADTKAPAKP  572

Qy    258  KEP--------------------APTTPKKPAPTTPKE-------PAPTTPKEPTPTTPKEPAPTTPKEPAPTTPK-EP  307
                :::                 |:|:|                   ::|
Db    573  RDPSPRKVAPTPAPEKKTPVLAKKEPAGPADSKTKEPEKSKPRDPSPRKKAVPAKPVRKTEV  632

Qy    308  APTAPKKPAP-------TTPKEPAPTTPKEP----SPTTPKEPAPTTTKSAPT  357
                |:|||             |::|                |
Db    633  APAAVKKPEPISKPKDTAPKKAEPNSPVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPE  691

Qy    358  TTTKEPAPTTTKSAPTTP-KEESPTTTKEPAPTTPK--EPAPTTPKKPAP--TTPKEPAPT  412
                |:|:|                  ::|               :|
Db    692  PEKK-TPVLAKKAPTKPDSEAAADPVSGPSSKDPKLAKKAPAVKPRDPSPMKAVPIKPAPK  750

Qy    413  TPKEPAPTTTKKPAPTA-PKEPAPTTPKETAPTTPKKLTTPTTP-----EKLAPTTPEKPA  466
                ||:|                  ::|                    :|
Db    751  T--EVPPAVVKKPEPVAKSRDPSPKKAK-AEPNSP--VVPPTPVKNPVKKKPPWEDDDA  805

Qy    467  PTTPEELAPTTPEEP------TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPK-PKE  520
                :::                 |::|                       ::|
Db    806  PAEPVNVPEPEKKTPVLAKKTPVKPRDPSPRKAVPAKPSTKTDAPPVSVKKPEPVSKPKE  865

Qy    521  PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPAPKELAPT-TTKEPTSTTSDKPAPTTP  579
                |:|                         ::|              :|
Db    866  PSPKRAEPNSPVVP---PTPVKNPVKKW-KPPWEDDDEPTEEVKPSE--PEKKTPVLA  918

Qy    580  KGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPT-----TPKKPAP------KEL  627
                |:|                                        :|
Db    919  K-KEPEKPKD-APKVAAKPRDSPRKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEA  974

Qy    628  APTTTKGPTS-------TTSDKP-----APTTPKET-------APTTKEPAPTTPKKPAPT  670
                |:|:            :::            :|
Db    975  PPTPVKNPVKKWKPPWEDDDEPAEPVSAPEPEKKTPVLAKKAPAKPRDP--SPKKAAPV  1031

Qy    671  TPETPPTTSEVSTPTTTKEPTTIHKSP----DESTPELSA-EP--TPKALENSPKEPG  722
                :::                          ::|
Db   1032  AAK-PDPKIPEV-PPTPVKNPVKWKPPWEDDDEPSEPVSAPEPEKKTPVLAKKAPTKPA  1089

Qy    723  V---------------PTTKITPAAT------KPEMTTTAKDKTTERDLRTTPETT--TAAPK  761
                                |:|:|           ||
Db   1090  TKPDSEAAADPVSGPTSKDPKLSKKAPAVEKRKPTTDEDKDDLKPSPAKKPEKAPEPAAPK  1149

Qy    762  MTKEFATTTEKTTESKITATTTQVSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTTE  821
                ||:|                                          |:|
Db   1150  KWKPVWDDDPDDEPEADFTVPAPSKKPDTEDPADPLG------GPKTKDPK---  1193

Qy    822  IMNKPEETAKPKDRAINSKATTPKPQKTKAPKKPTSTKKRKTMPRVRKPK--------  872
                ||:|                            ::|
Db   1194  -LNKRAPAEKPTEK----------PKPKEVSKEPPKPTEPPKP-AAPKKWKPPWEDDDEPE  1243

Qy    873  ---TTPTPRKMTMTSTMP  885
                   :|:|
Db   1244  ADFTMPAPKKPDTEDP  1259
```

RESULT 11

Q9LIE8

ID   Q9LIE8   PRELIMINARY;   PRT;   1480 AA.

```
Qy     30  KNRTKK-KP----------KPPVVDEAGSGLDNGD---FKVTTPDTSTTQHNKVSTS  74
                || ||  |              :|                 :|||
Db    274  KNPTKKWKPPWEDETPVEEVKEPPVPEKKAPVLKKKDPAPAAKARDPSPSKAAPKKVEPS  333
```

Query Match   12.2%;   Score 802;   DB 5;   Length 1274;
Best Local Similarity 28.0%;   Pred. No. 1.2e-45;
Matches 290;   Conservative 102;   Mismatches 414;   Indels 230;   Gaps 56;

```
AC    Q9LIE8;
DT    01-OCT-2000 (TrEMBLrel. 15, Created)
DT    01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS    Arabidopsis thaliana (Mouse-ear cress).
OC    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC    eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX    NCBI_TaxID=3702;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=COLUMBIA;
RA    Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL    Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=COLUMBIA;
RX    PubMed=10907853;
RA    Nakamura Y.;
RT    "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT    sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT    TAC and BAC clones.";
RL    DNA Res. 7:217-221(2000).
DR    EMBL; AP001306; BAB03062.1; -.
DR    InterPro; IPR002965; P_rich_extensn.
PRINTS; PR01217; PRICHEXTENSN.
SQ    SEQUENCE 1480 AA; 147153 MW; D1AC0C79F155E732 CRC64;

Query Match    12.2%;
Best Local Similarity   29.6%;   Pred. No. 1.7e-45;
Matches  307;  Conservative  54;  Mismatches  505;  Indels  171;  Gaps  45;

Qy     34 KKKPKPKPKPPVDEAGSGLDNGDFKVTTPDISTSTQHNKVSTS----PKITTAKPINPRPSL  89
          ||: |||| |||                        :|:|     |:
Db    117 KPRPHPKPDNVK-------------PHPHPKPTPKHPHPHPKPTPKHHHPHPKPPTPKHPKPPSV 167

Qy     90 PPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPT 149
          |||                   ||:
Db    168 KPPPSTPK------------PPTTNPPPSTPQPPPHKPPCPTPTPPVASPPMATPP 211

Qy    150 SKV--LAKPTPKAETTTKGPALTTPKEPPTP---TTP-KEPASTTPKEPPTTTIKSAPTTP 203
          :::  |:| :||| |||||||| ||||||       |  | ||||||| |||| ||||||
Db    212 TQMPIATP-PIAKSPVATPPIATPPITTIPPVATPPITTIPIANPPIMPPIATP 270

Qy    204 KEPAPTTT------------KSAPTTPKEPAPTTT--------------------KEPAPTTPKEPAPTTTKE 244
          ||||                ||||                              ||||
Db    271 PVAAPPITNPPISKPPVTTP---PTTTPPPIAKPPIATPPISTPPAATTPAATPPITTLP 326

Qy    245 PA-------PTTTKSAPTTP-KEPAPTTPKPAP--TTPKEPAPTTPKEPTPTTPKEPAP 294
          |:      :|| |||| | ||||||     || |||||||| |||
Db    327 PAKPPVAISPIVT--PPVTPIAQPPVATPPTATPPVATPPIATPPTSKSPISTPPISESP 384

Qy    295 TTTKEPAPTTP-KEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTTI----KEPSPTTPKEPAP 349
          ||||||||||  ||||                               |:|| ||||||||
Db    385 VATPPTATSPIKTPPPAKPPVATPPIAKSPIATPPTATPPVATPPIEKPPVATPPTTTPP 444

Qy    350 TTTKSAPTTTKEPAPTTTKSAPTTPKEP---SPTTTKEPAPTTPKEPAP----TTPKPAPT 404
          |||| ||||| |||||| |||| ||||   ||||| |||||||| |||    |||| |||
Db    445 TAT---PPVAKPVETPPIATPPPATPPIAKPPISTPPISKPVVATPPAATPPITTTPVKPPVA 501

Qy    405 TPKEPAPTTPKEP---APTTTTKKPAPTAP---KEPAPTTPKETAPTTT------PKKLTP---T 452
          |||||||||||||   :||||  ||||||||   ||||||||| ||||       :
Db    502 TPPLAIPPVAKPPVVTPPATPPIATPPVATPPIAKPPVTPPTT 561

Qy    453 TPKELAPTTPEKLAPTTPEEPTTPEELAPTTT-PEEPTTTPEEPAPTTPKAAAPNT--PKEPAPT 508
          |||                        |||    |: ||||| |||||| |:
Db    562 TPPIATPPVAKPPVATPPIATPPIATPPPISTPPISKSPVATPPAATPPIITTTPPPAKPPVA 621

Qy    509 TPKEPAPTTPKEPAPTTPKKPAPKELAPTTT 563
          ||:              :| :|||
Db    622 TPPIATPPIAKPPVATPPATPPIATSPVAKPVVATPPIKTPPPAKPPVAIPPIATPPVA 681
```

```
Qy    564 KEPTSTTSDKPAPTTPKGTAP--TTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKP 622
          ||:|    :||| ||          |||||||| ||||||| |||       ||:|| ||
Db    682 KPPVAT---PPTATPPIATPPIATPPVVTPPIATPPVATPPIAKPPTTI--PPTATPPVA 736

Qy    623 APKELAPTTTKGPTST---TSDKPAPTTPKETAPT-TPKEPAPTTPKPAPTTP-----E 673
          ||                         |||| |||   |:  |||||| |||      |
Db    737 MPPIATPPTAKPPIATPPIAIPPVAKPPVVTPPTATPPIAIPPIAKSPVATPPTATPPVA 796

Qy    674 TPPPTTSEVSTPTTTTKEPTTIH--KSPDESTPELSAEPTPKPALENSPKEPGVPTTKPPAA 731
          |||| |     |||| |||||||  |                 |:
Db    797 TPPIAKPPVATPPTTTAPPTAATPDVAKPVATPPIATPPTAKPPILTPPISKPPVATPPAA 856

Qy    732 TKPEMT-TTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTT 790
          ||:|   |:| ||: |||||     |:|   ::|  |||| |||:
Db    857 TPPITTPPPAKPPVATPPIATPP--IAKPPVATPPIATPPIATSPVAKPPVAIPPIKTPP 914

Qy    791 QDTTPFKITTLKTTTTLAPKVTTTTKKT---IITTEIMNKPEETAKPKDRAINSKATTPKP 846
          | |||   |::||:|||||||||||       |:||| |||||| || ||| ||||| ||
Db    915 PAKPPVAIPPIATPPVAKPPVATPPTATPPIATPPIATSPIATPPVVT---PPTATSPVATPPIA 971

Qy    847 QKPTKAPKKPISTKKPKTMPRVRKPKTTPTRKMTSTMPELNP--------TSRIAE--- 895
          :|:| | | |:|| | |  | |||||||                    :|| |
Db    972 KPPTTTP--PTAT-PPVAMPPIATP---PTAKPPVATPPIANPPVEKPPVATPPIAKPPT 1025

Qy    896 ----AMLQTTTRPNQTPNSKL--VEVNPKSEDAGGAEGEETPHMLLRPHVFMPEVTPDMD 948
                  |||||||          || |   ||  |  |        |:|:
Db   1026 VLPPIAKPPVETSPTATPPIATPPVAIPP-----------VVKPPVAIPPIKPPV 1070

Qy    949 YLPRVPNQGIIINPMLS 965
          ||||        |  |:
Db   1071 ATPPVTNPPTAMPPIVT 1087
```

```
RESULT    12
ID    Q41805   PRELIMINARY;    PRT;   1188 AA.
Q41805
AC    Q41805;
DT    01-NOV-1996 (TrEMBLrel. 01, Created)
DT    01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    EXTENSIN-LIKE PROTEIN PRECURSOR.
OS    Zea mays (Maize).
OC    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC    Panicoideae; Andropogoneae; Zea.
OX    NCBI_TaxID=4577;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=B73; TISSUE=POLLEN;
RA    Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL    Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; Z34465; CAA84230.1; -.
DR    Mendel; 14346; Zeama; 2368; 14346.
DR    InterPro; IPR001611; LRR.
DR    InterPro; IPR003592; LRR_out.
DR    Pfam; PF00560; LRR; 3.
DR    SMART; SM00370; LRR; 4.
FT    SIGNAL.
FT    SIGNAL     1     27     POTENTIAL.
SQ    SEQUENCE 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

Query Match    11.9%;
Best Local Similarity   29.9%;   Pred. No. 1.7e-44;
Matches  230;  Conservative  59;  Mismatches  359;  Indels  122;  Gaps  33;

Qy     37 PTPKPPVVDEAGSGLDNGDFKVTTPDISTTQHNKVSTSPKITTAKPINPRPSL----PPN 92
          ||         ||||||||||| |   |   ||||| |      | ||||||    ||:
Db    460 PTPHSPPAD------DYVPPTPVPGKSPPATSPSPQVQPPAASTPPSLVKLSPPQ 510

Qy     37 PTPKPPVVDEAGSGLDNGDFKVTTPDISTTQHNKVSTSPKITTAKPINPRPSL----PPN 92
          |||
Db     93 SDITSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV 152
```

```
Db  511 APVG--------------------SPPPPVKTSPPAPIG---- 541
             :  :                :: : : :
Qy  153 -----LAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTKEPA 207

Db  542 SPPPPVKSPPPPAVGSPPPPEKSPPPAPIT--PPPPLVASPPPPVKSPP 599
             : :                           :   :
Qy  208 PTTTKSAPTTP--KEPAPTTPKEPAPTTTKEPAPTTTKSAPTTP-KEPAPTT 264

Db  600 PPAPVASPPPPVKSPPPPTPVASPPPPAVPASSPPPPAPVASPPPPEKSPPPP 659
            : :         :            ::      :
Qy  265 PKKPAPTTPKEPAPTT-----PKE--PTPTTPKEPAPTTPKEPAPTTAPKKP 315

Db  660 PAKSTPPPPEEYPPPTSVKSSPPPEKSLPPPLLIPSPPPQEKPTPPSTPKEPAPTTTKEPAPTTTKSAPTTP-PSSPEKP 718
                      :       ::   :  :        :  :   :::: :  ::::  :::
Qy  316 APTTPKEPAPTTPKEPAPTTPKEPAPTTTKEPSPTTTKEPAPTTKSAPTTTKEPAPTTTKSAPTTP 375

Db  719 SP--PKEPVSSPPQTP-----KSSPPPAPVSSPPPTPVSSSPPPALAPVSSPPPSVKSSP---- 768
            :: :                :    ::       :
Qy  376 EPSPTTTKEPAPTTPKEPAPTTPKKPAPTTPKKPAPTTTKKPAPTTPKEPAP 435

Db  769 PPAPLSSPPPAPQVKSSPPPVQVSSP--PPAPKSSPPLAPVSSPPPQVEKTPSPPAPLSSPP 827
            :                  ::                :
Qy  436 TTPKETAP---------TTP---KKLTPTTPKLAPTTPEEKP-----TTPEELAPTTPEEPT 482

Db  828 LAPKSSPPPHVVVSSPPPPVVKSSPPPSPPLTPKPASPPAHVSSPPPEVVKPSTP--PA 885
            :                 ::                   :   :
Qy  483 PTTPEEPAPTTPKAAAPNTPKE-PAPTTPKEPAPTTPKETAPTTPKGTAPTT 541

Db  886 PTTVISP--PSEPKSSPPPIVKSSPPPAMVSSP-PMTPKSSPPPVVSSPPPT 943
            ::                  ::           :
Qy  542 LKE-------PA-----PTTPKK---PAPKELAPTTTKE---PTSTTSDKPAPTTPKGTAPTTP 587

Db  944 VKSSPPPPAPVSSPPATPKSSPPPAPVNLPPPEVKSSPPPSPPPVSSPPPA---PKSSPPPAP 1000
            :                          :              ::
Qy  706 SAEPTTPKALENSPKEPGVPTTKTPAAATKPEMTTTAKDKTTTERDLRTTPET 755

Db 1058 KSPPPPAPISSPPPPVKSPPPPAPVS--SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS 1115
            :                         :
Qy  647 -TPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTTKEPTTIIHKSPDESTPEL 705

Db 1001 MSSPPPEVKSPPPPAVSSPPPPAPVSSPPPPPV---PPVKSPPPPAPVSSPPPPV 1057
            :
Qy  588 -KEPAPTTPKEPAPTTPKGTAPTTLIKEPAPTTLKEPAPTTPKKPAPTTKGPTSTTSDKPAPTT 646

Db 1116 SPPPAP-----VKPPSLP-PPAPVSSPPPPVVTPAPPKEEQSLPPPAES 1158
            ::                 :
Qy 1058 -----TTTKKPAPTTAPKEPAPTTPKETAPTTPKKLTLPTTPKEKLAPTTPEEK-PAPTTPEE 472
```

```
Query Match     11.9%;  Score 778.5;  DB 11;  Length 2187;
Best Local Similarity 27.9%;  Pred. No. 7.8e-44;
Matches  292;  Conservative 119;  Mismatches 420;  Indels 215;  Gaps 48;

Qy   22 VSSVKDNKNKNRTKKPTRKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNK----VSTSPK 76
            : ::         :                              :    ::
Db  803 ISAVQSPKVDPIMSDVTPTSPKKTSA--------------TAVPKDTSATLSLKSVPAVTSLSPP 853

Qy   77 ITTAKPINPRPSLPPNSDDISKETSL--TVNKET-----TVEIRKETTTTNKQTSTDGKEK 128
            :  :                  :: :         :
Db  854 KAPVAPSNEATIVPTEIPTSLKNALAAATPKETLATSIPKVTSPSPKTPKSVSLKGAPA 913

Qy  129 TTSAKETQSIEKTISAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTP 188
            :      :                       :   ::::
Db  914 MTSKKAT--EIAASKDVSPSQ--FPKEVPLLQHV--PPTSPPKSPVSDTLSGALTSPP 965

Qy  189 KEPTPTTIKSAPTTPKE-PAPTTTKSAPTTPK-----------EPAPTTTKEPAPTTPKEPA 238
            :        :       :                         :
Db  966 PKGPPATLAETPTVPKKSPKPAASKKTPATPSPEGVTAVPLEIPPCSKKAPKTAAPKESS 1025

Qy  239 PTTTKEPAPTT--TKSAPT----------------------TPKEPAPTTPKK-PAPTTPK 274
            :                                        :
Db 1026 ATSSSKRAPKTAVSKEIPSKGVTAVPLEISLPLKETSKSATPGEKSASSPKRSPKTAGPK 1085

Qy  275 EPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKE-PAPTTTKEPAP----TTP 328
            :                                 :
Db 1203 GEISPSPKKTSKTAAPKENSATLPPKRSPKTAAPKETPATSSEGVTAVPSEISPSPPTPA 1262

Qy  376 -EPSPTTTKEPAPTTPKEPAPTTPKKPAPTTP--PKE----------PAPTTPKEPAP- 419
            :                                 :
Db 1143 LEIPSAPQKARAPKTAVPKQIPTPEDAVTILAGSPLSPPKKASKTAAPKEAPATPSVGVIAVS 1202

Qy  329 KEPAPTTTKEPSPTTPK-----EPAPTTTKSAPTTPKEPAPTTT-KSAPTTPK------- 375
            :                     :
Db 1086 E---TPPGGVTAVPPEISLPPKETPQNATPNESLAASSQKRSPKTSVPKETPPGGVTAMP 1142

Qy  473 LAPTTPPEEPTPPTT---------PEEPAPTTPKAAAPNTPKEPAPTTPKE-PAPTTPKEPAP 523
            :                        :
Db 1316 ASATPPSKKTPKTAVPKETSAPSEGVTAVPLEIPPSPRKAPKTAAPKETPAPS--PEGAT 1373

Qy  524 TTPKETAPTTPKGTAPTTLIKEPAPTTP----------------------KKRAPKELAPTTT 563
            :                                                  :
Db 1374 TAPVQIPPSPRKGSKKAGSKE-TPTTPSPEGVTAAPLEIPISSKKTSKMASPKEPLVTPS 1432

Qy  564 KEPTS-----------TTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPAP 616
            :
Db 1491 SSPKK-APKTAAPPSER--VTTVPPEKPA-TPQKASGTTASKVPVPRAETQEVAVSSRETPV 1547

Qy  617 TTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKEPAPTTPETP- 675
            :
Db 1433 SKKLSQTVGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLTPSPK-DAPTTLAE-SP 1490

Qy  676 ----PPTTSEVSTPTTTK------EPTTIHKSPDESTPELSAEPTPKALENSPKE-PGVP 724
            :
Db 1548 TPAVPVKNPSSHKKTKTSKTIELKEEAPATLPPSTTKSPKIPSSKKAPRT--SAPKEPPASP 1605

Qy  725 TTKTPAAATKPEMTTTAKDKTTERDLRTTPETTTAAPKKMTKETATTTEKTTTESKITATTTQ 784
            :
Db 1606 SIK-PVTT--SLAQTAPPSLQKAPSTTIPKENLAAPAV-----LPVSSKSPAAPARASASL 1658

Qy  785 VTSTTTTQDTTPFKITTLKTTILAPKVTTTIKKITTTTEIMNKPEETAKPKDRATNSKATTP 844
            :
Db 1659 SPATAAPQTAAPKEATLIPSCKKAAATETPIETSTAPSLEGAPKETSE----TSVSKVLMS 1714
```

```
Qy    845 KPQKPIKAPKPTSTKKPTMPRVRKPKTTTPPRKMTSTMPELNPTSRIAEAMLQTTTRP 904
          |||:|:| |: |:|             | |     |  ::  :|
Db   1715 SP------PKRASSSKRASTLP---------------ATTLPSLKEASVLS--------P 1745

Qy    905 NQTPNSKLIVEVNPKSEDAGGAEGETP 930
           ||  |  |  |:|  ::||   ||
Db   1746 TATSSGKDSHISPVS-DACSTGTTTP 1770
```

RESULT   14
Q9VEL9

ID   Q9VEL9   PRELIMINARY;   PRT;   2112 AA.
AC   Q9VEL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG4090 PROTEIN.
GN   CG4090.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003716; AAF55402.1; -.
DR   FlyBase; FBgn0038492; CG4090.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF01607; Chitin_bind_2; 10.
DR   SMART; SM00494; ChtBD2; 11.
DR   PROSITE; PS00022; EGF_1; 1.
SQ   SEQUENCE 2112 AA; 219547 MW; B91018E551A5D36 CRC64;

Query Match   11.3%;
Best Local Similarity 27.2%;   Score 739; DB 5; Length 2112;
                       27.2%; Pred. No. 3.3e-41;
Matches   281; Conservative 98; Mismatches 339; Indels 316; Gaps 44;

```
Qy     28 NKKNRTKKKPTKPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRP 87
          |: |                  :                  :|  : : ::
Db   1023 NHPDQIQTKPLRCKKVVSQGGSSSN-----------STSNSSSSSSNNSGSGSSSNGSGSSS 1070

Qy     88 SLPPNSDTSKETSLTVNKETTV--ETKETTTTNKQTS----TDGKEKTTSAKETQSI-- 138
           | | ||| | ||     ||       ||:|    |:       | |  : || |:
Db   1071 SSSSNSGGSSSNTGSSSNSGASSSGGGSNQGSSSNSGSGSSSNSSGNQSTSSSTSSSSSSS 1130

Qy    139 ----EKTSAKDLAPTSKVLAKPTKPKAETTTKGP------------------- 167
              :| :||:|||| | |    || |:|||| |
Db   1131 NNNNQGSSSSSSSSSSSTSSKPNPSEFCKVNGQFIGDRSDCAKFYRCVDNDRGGENMVPF 1190

Qy    168 ----------ALTTPKEPTPTTPKEPASTTPKEPPTTPTTIKSAPTTT 202
                      | ||||| || | |||  |||  | |   | |||||
Db   1191 SCGPGTVWDAQMQACNHAWAVKECGGIAPPTTSTPTTSR-PTTASTSRPSDQTTSTSRPTG 1249

Qy    203 PKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTT---- 241
          | |  || | |  |  |  | ||
Db   1250 P-----PTTAREVTIARPTTSSPTTASSSQTTSPVTQAPNTDGKCRSEGFMADPNNCSKFYR 1305

Qy    242 -----------TKEP-----APT------------TTKSAPTTPKEPAPTTPKKPAPTT 272
                      ||||        ||             || |||||| |||| |  | |
Db   1306 CVRNNKGGFTSIPFCQCGAGTVWDQDLQTCNHNFNNCSTGTESTTPKPPC--EPATNGTTA 1363

Qy    273 PKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 332
          | || | |     || |  | |:| :   |
Db   1364 TSTSSTTTP--PPTTTDLPPTPSTTGLP-PTTTTELPPT------TTTDLPPTTTTRLPP 1413

Qy    333 PTTTKEPSPTTPKEPAPTTTIKSAPTTPKEPAPTTTPKSAPTTPK--EPSPTTTKEPAP-- 387
          |||   | || |  | || |   | |  | || |: |  |||| |   |  ||| |||
Db   1414 TTTTSLPPTTTTGLPPTTTGAQPTTTTLSSETETSTVTISPESTTQPPSTTTMKPLPAG 1473

Qy    388 -------------------------------TTPK-------------EPAP---TT 397
                                         || |              |||   ||
Db   1474 TECTGGEGYMADPEDCRKYYRCINAGASYRKYNFTCPKGTGWNEEVQTCDTVENIPRCSKL 1533

Qy    398 PKKPAPTTP----KEPAPTTPKE-PAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKKLTPT 452
          |: | || |    ||| ||| |  | ||| : ||   | |||| |  | |   :   | |
Db   1534 PAEPITTTPSEESKDPGSTTPQSTDEPTTVTKPI-TKPTE-EPSTEKPOKPTTTQYPEKPT 1591

Qy    453 TPEKLAPTTEPKPAPTTPEELAPTTPEEPTTPEEPAPTTPKAAAPNTPKEPAPTTPKE 512
          |  : |  | || | |   | || |    |    | |       | | |  |  ||
Db   1592 TTEE--PEKPQKPT-TTEYPQKPTTTEEPPPEKPQKPT-TTEYPQKPTTTEEPPPEKPQK 1647

Qy    513 PAPTT-PKEPAPTTPKETA-----PPTPK--GTAPTTLKEPAPTTPKKPAPKELAPTTTK 564
          | |   |  | ||  |        | | |   | |  ||| ||| || | |  |
Db   1648 PTTTEEYPQKPTTEEPTTSILPGYNPTTTSVPGYNPTTTPIPVETTTSTPGYK---PTTTG 1704

Qy    565 EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKK-- 621
          || : |  ||| ||||  | || | |  | |    | ||||  |  ||||| || |
Db   1705 EPITTTT-LPSTTTDAIQEPPTSKKPEPTTTTESPESSTPEGSVTTLQPEPQPNYNCSSE 1763

Qy    622 ---PAPKELA-----PTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTP 672
             || |       || | |  | |||||| |||  ||  ||  | || |  | || |
Db   1764 GFFPDPEDCSRYYRCVDAAKNGKYQVYAFKCGKGTVWDTSTETCNYADQVSGN------ 1816

Qy    673 ETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAT 732
          :  : || |   | || |  |
Db   1817 ----CSSGQTTTPGTTTEPGTTESTTSSGKP---------------------TTTWA-- 1857

Qy    733 KPEMTTTAKDTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQD 792
          || ||||: :||        || ||   |                    ||: ||| |
Db   1858 -PETTTT-------------SSPETTTT---VASETTTTTSGT------TTTATPETTTKP 1895

Qy    793 TTPFKITTLKTTTLLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTTRPKQKPTKA 852
          |||  |||||||:   |  ||||            |  |          |||||
Db   1896 PKP-----EITTIAGEETSTSKSPTTTE--------------------SPAPSTNTSA 1928
```

```
Qy    853 PKKPTPTSTKKPKTMP 866
              |:|:|
Db   1929 P------CPETGP 1935
```

```
RESULT   15
Q9XDH2
ID    Q9XDH2;                   PRELIMINARY;       PRT;       763 AA.
AC    Q9XDH2;
DT    01-NOV-1999 (TrEMBLrel. 12, Created)
DT    01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    PROLINE-RICH MUCIN HOMOLOG.
OS    Mycobacterium tuberculosis.
OC    Bacteria; Firmicutes; Actinobacteridae;
OC    Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX    NCBI_TaxID=1773;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=H37RV;
RA    Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA    Zhang Y., Moreno C., Singh M.;
RT    "Cloning and characterization of a new member of the PGRS family that
RT    is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL    Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
EMBL; AF071081; AAD41594.1; -.
InterPro; IPR002951; Atrophin.
InterPro; IPR002965; P_rich_extensn.
InterPro; IPR003882; Pistil_extensin.
PRINTS; PR01222; ATROPHIN.
PRINTS; PR01217; PRICHEXTENSN.
PRINTS; PR01218; PSTLEXTENSIN.
SEQUENCE  763 AA;   75034 MW;   39168EC45A5916F8 CRC64;
```

```
Qy    612 KEPAPTTP----KKPAPKELAPTTTKGP-TSTTSDKPAPTTPKET---APTTPKEPAPTTP 664
              ||:|:|          ||  || ||  |   |  |  || |       ||| ||
Db    489 APPAPPSMPSAVRVPPSPPIPAPPAAPRASMPALPPAPPSPPPATRLCPPLPPSPPAPNSP 548
```

```
Qy    665 KKPAPTTPEETPPPTTSEVSTPTTTKEPTTIIHKSP-----DESTPELSAEPTPKA--LENSP 718
              ||| |||                |  |   |  |:|   | |||:|||| ||     ||||
Db    549 --PAPPAPPTTPKLLS--ANPCPPVPPAPNRPPAPPAPPAPPAPPELPAPPDDPPTPVANSP 604
```

```
Qy    719 KEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKI 778
              | | | || |  | || ||||         || ||         |  || ||   |
Db    605 --PAPPAPPAAPPSALPFVNPPA------------PPTPAAPK-------------SRPAL 637
```

```
Qy    779 TATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTTITTTEIMNKPEETAKPKDRATN 838
              |                           |              |            |
Db    638 PAAPPAPPAAPPVRATTP-------------------------------PPAPPAPPAPN 665
```

```
Qy    839 SKATTPKPQKP-----TKAPKKPTSTRKPKTMPRVRKPKTTPTTPRKMTSTM----PELNP 889
              | | | ||         |  |:|     | |                  |    | |
Db    666 SMALPPAPPDDPPIPLLATPPAPPAPPLPMSPPAPPLPPAAPDPPAPPLIINQPPSPPLAP 725
```

```
Qy    890 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET 929
              : |:  || |                | |:|
Db    726 VPGAPLAPLPINGRPVFARKNSLI--------GSSSGDT 756
```

```
Search completed: April 26, 2002, 16:28:53
Job time: 548 sec
```

```
Query Match    10.9%;    Score 715;    DB 2;    Length 763;
Best Local Similarity   28.2%;    Pred. No. 4.6e-40;
Matches  248;  Conservative  39;  Mismatches 361;  Indels  232;  Gaps  39;
```

```
Qy    156 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 215
              | |   |  |           || ||       ||| ||        :|
Db      3 PVP-------APRALAPLPPAPPAPPAEPKSKPPPPPAP-------PAPPCWMLVSAAP 46
```

```
Qy    216 TTTPKEPAPTTPKPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKKPAPTTTPKKPAPTTPKE 275
              ||| ||||||                ||||   || |||       |: |||
Db     47 PCP--PAP----PAPPKPKSKAPFPPPVPPAPPARELAPPLP-PAP-------PEAPRE 90
```

```
Qy    276 PAPTTPKEPTPTT-----PKEPAPTTKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK 329
              |       |  |       ||  |  | |    |  ||  | |  || | |||  || |
Db     91 SRPALPPCPPPPVVIPDDPEPAAPPVPPAPNSPFPPPPFPPAPKFVPAPPVP--PVPNSPP 148
```

```
Qy    330 EP--APTTTKEPSPTTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAP 387
              ||   ||| || ||||||     ||   || |||||    ||  ||| |  || |||||||
Db    149 FPFPPPPAALNPPAP-----PAPPLANSPLPPPAPPTPAGT--PPAAPWPPVPAAPKSKPA 201
```

```
Qy    388 TTTPKEPAPTTPKKPAPTTPKEPAPTTTPKEPAP----TTTKKPAPTAPKEPAPTTPKETAP 443
              ::|:|||||    |||| |||||  | |||:||       | |:||| |||| ||| |
Db    202 SPPRPPAP----PMPATPMEFPPLPVPPDPISKEFPPAPPAPPIPPAPVPIPPVPPLP 256
```

```
Qy    444 TTTPKKLTPTTPEKLAP-------------TTTPEKPAPTTPEELAPTTPEEPTTTPEEPAPTTP-- 493
              |  | | |  |  ||                  |                |      ||:  |
Db    257 PVPNKIPPAPP---APPVAVAAVLVAQCPPLPPLPNNHPAPPAAAPVPGVPLAPLPNSHP 313
```

```
Qy    494 ---PKAAAPNTPKEPAP------------------------------------------507
              |
Db    314 PAPPSAPVPGVPLAPLPISGRPVSVWKGSFTTLSTFCCRVCSGEVLAGALNPSRPSRSPL 373
```

```
Qy    508 --TTPKEPAPTTPKEPAPTTPKETA----PTTPKGTA---PTTLKEPAPTTP--KKPAPKE 557
              |||||                      |||          |:  |||| |  || ||
Db    374 TTTTPALPAPIPPLPLPPLPLPINTAVPPIPPLPPVTALAPLLPPLAPLPISPGVPPAP-- 431
```

```
Qy    558 LAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPA------PTTPKEPAPTTPKGTAPTTL 611
```

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: April 26, 2002, 16:29:32 ; Search time 210.7 Seconds
(without alignments)
948.306 Million cell updates/sec

Title: AA6
Perfect score: 7294
Sequence: 1 MAWKTLPIYLLLLLSVFVIQ.........ARAITTRSGQTLSKWYNCP 1366

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB Match	ID	Description
1	7265	99.6	1404	4	Q92954	homo sapien
2	7236	99.2	1404	4	Q9BX49	homo sapien
3	3759	51.5	1054	11	Q9JM99	mus musculu
4	1713.5	23.5	401	6	O77765	bos taurus
5	982.5	13.5	1079	5	Q9N4S7	caenorhabdi
6	946.5	13.0	1049	5	Q9ITS1	drosophila
7	891.5	12.2	1795	5	O76894	drosophila
8	863.5	11.8	1489	10	O96649	phytophthor
9	846	11.6	1274	5	O20007	caenorhabdi
10	817	11.2	1315	10	Q9SPM0	zea mays (m
11	814	11.2	1480	10	Q9LIE8	arabidopsis
12	812	11.1	2187	11	P70670	mus musculu
13	784	10.7	1188	10	Q41805	zea mays (m
14	781	10.7	2112	5	Q9VEL9	drosophila
15	715	9.8	763	2	Q9XDH2	drosophila
16	707.5	9.7	555	10	Q9FPG6	mycobacteri
17	706.5	9.7	6677	10	Q9N435	chlamydomon
18	703.5	9.6	4880	11	Q9ILT1	rattus norv
19	703.5	9.6	5085	11	Q9JKS6	rattus norv
20	695	9.5	1151	13	O57580	gallus gall
21	680	9.3	1229	5	Q94185	caenorhabdi
22	667.5	9.2	2284	5	Q9VPG1	drosophila
23	665	9.1	4833	11	Q9QYX6	mus musculu
24	665	9.1	5038	11	Q9QYX6	mus musculu
25	664	9.1	2089	4	Q9QYX7	homo sapien
26	661.5	9.1	7962	4	O14676	homo sapien
27	660.5	9.1	2768	5	O10465	homo sapien
28	653.5	9.0	1514	5	Q9CC00	drosophila
29	649	8.9	3570	4	Q9GUW7	leishmania
30	647.5	8.9	3507	10	Q99552	homo sapien
31	643.5	8.8	990	13	Q91803	xenopus lae
32	636	8.7	6632	5	O17362	caenorhabdi
33	633	8.7	489	10	Q17707	vigna ungui
34	632	8.7	761	10	Q92Q10	arabidopsis
35	631	8.6	2344	5	Q9N3Y8	caenorhabdi
36	629	8.6	1612	5	O01761	caenorhabdi
37	628.5	8.6	971	5	Q9VYQ2	drosophila
38	627.5	8.6	839	2	Q9XVS4	caenorhabdi
39	622.5	8.5	801	5	Q9RX57	deinococcus
40	607.5	8.3	801	5	Q9XVS7	deinococcus
41	607.5	8.3	924	12	Q23635	caenorhabdi
42	605	8.3	379	5	Q99307	epstein-bar
43	600.5	8.2	379	5	Q27929	drosophila
44	598.5	8.2	1893	5	Q9NKC9	drosophila
45	592.5	8.1	2232	5	Q9SBM1	volvox cart
					P91365	caenorhabdi

ALIGNMENTS

RESULT 1
Q92954

ID   Q92954   PRELIMINARY;   PRT;   1404 AA.
AC   Q92954;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEGAKARYOCYTE STIMULATING FACTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J.; Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Blood 78:279-279(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Merberg D.M., Fitz L.J.; Temple P., Giannotti J., Murtha P.,
RA   FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA   Jacobs K., Turner K.;
RL   (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL   Mosher D.F. (eds.);
RA   Biology of vitronectins and their receptors., pp.45-52,
RL   Elsevier Science Publishers B.V. (1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J.; Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Blood 78:279-279(1991).
DR   EMBL; U70136; AAB09089.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   InterPro; IPR024001; Somatomedin_B.
DR   Pfam; PF00045; Hemopexin; 2.
DR   Pfam; PF00086; Somatomedin_B; 2.
DR   Pfam; PF01033; hemopexin; 2.
DR   PRINTS; PR00438; GFCYSKNOT.

```
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
SQ   SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match            99.6%;   Score 7265;  DB 4;  Length 1404;
Best Local Similarity  97.3%;   Pred. No. 0;
Matches 1366;  Conservative  0;  Mismatches  0;  Indels 38;  Gaps 1;

Qy    1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRGEGYSRDATCNCDYNCQHYMECCPDF  60
        ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRGCEGYSRDATCNCDYNCQHYMECCPDF  60

Qy   61 KRVCTA------------------------------------AEVHNPTSPSSKKAP  82
        ||||||                                    |||||||||||||||
Db   61 KRVCTAELSCKGRGCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

Qy   83 PPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSTIW 142
        |||||||||||||||||||||||| ||||||||||||||||||| ||||||||||| ||
Db  121 PPSGASQTIKSTTKRSPKPPNKKTKTKKVIESEEITEEHSVSENQOESSSSSSSSSSSSTIW 180

Qy  143 KIKSSKNSAANRELOKKLKYKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST 202
        |||||||||||| ||||| |||||||||||||||| ||||||||||||||||||| ||||
Db  181 KIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTPKRPVVDEAGSGLDNGDFKVITPDTST 240

Qy  203 TOHNKVSTSPKITTAKPINPRPSLPPNSDDTSKETSLITVNKETTVETKETTTTNKQTSTDG 262
        | ||||||||||||||||||||||||||| ||||||| |||||||||||||||||| |||
Db  241 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTIVNKETTVETKETTTTNKQISTDG 300

Qy  263 KEKTTSAKETQSIEKTSAKDLAPISKVLAKPTKAETTTKGPALTTPKEPTPTTPKEPAS 322
        |||||||||||||||||| ||| |||||||| ||||||||||||||||||||||||||
Db  301 KEKTTSAKETQSIEKTSAKOLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS 360

Qy  323 TTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKEP 382
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKEP 420

Qy  383 APTTTKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 442
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 APTTTKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 480

Qy  443 EPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKSAPTTTKEPAPT 502
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 EPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKSAPTTTKEPAPT 540

Qy  503 TTKSAPTTPKEPSPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKKP 562
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TTKSAPTTPKEPSPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKKP 600

Qy  563 APTAPKEPAPTTPKETAPTTPKKLIPTTPEKLAPTTPEELAPTTPEEPTPTT 622
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 APTAPKEPAPTTPKETAPTTPKKLIPTTPEKLAPTTPEELAPTTPEEPTPTT 660

Qy  623 PEEPAPTTPKAAAANTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKRGTAPTTLKEP 682
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 PEEPAPTTPKAAAANTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKRGTAPTTLKEP 720

Qy  683 APTTPKKPAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKG 742
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 APTTPKKPAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKG 780

Qy  743 TAPTTLKEPAPTTPKPKPAPKELAPTTTKGPTSTTSDKPAPTTPKEPAPTTPKEPAPTTPK 802
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 TAPTTLKEPAPTTPKPKPAPKELAPTTTKGPTSTTSDKPAPTTPKEPAPTTPKEPAPTTPK 840

Qy  803 KPAPTTPETPPTTPTTISEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT 862
        |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  841 KPAPTTPETPPTTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT 900

Qy  863 TKTPAATKPEMTTPAKDTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 922
        |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  901 TKTPAATKPEMTTPAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 960
```

```
Db  901 TKTPAATKPEMTTPAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 960
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  923 TSTTTQDTTPFKITTLKTTTLAPKVTTTKKITTTTEIMNKPEETAKPDRATNSKAATTPK 982
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 TSTTTQDTTPFKITTLKTTTLAPKVTTTKKITTTTEIMNKPEETAKPDRATNSKAATTPK 1020

Qy  983 PQKPTKAPKPTSTKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTPRPN 1042
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 PQKPTKAPKPTSTKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTPRPN 1080

Qy 1043 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1102
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1140

Qy 1103 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPPSPARRITEVWGIPSPIDTVFT 1162
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPPSPARRITEVWGIPSPIDTVFT 1200

Qy 1163 RCNCEGRKTFFFRDSQYWRFTNDIKDAGYPKPIFKGFQGGLTGQIVAALSTAKYKNWPESVY 1222
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 RCNCEGRKTFFFRDSQYWRFTNDIKDAGYPKPIFKGFQGGLTGQIVAALSTAKYKNWPESVY 1260

Qy 1223 FFKRGGSIOQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1282
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 FFKRGGSIOQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1320

Qy 1283 SPARLAYQDKGVLHNEVKVSILWRGLPNVVISAISLPNIRKPDGYDYYAFSKDQYNIDV 1342
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 SPARLAYQDKGVLHNEVKVSILWRGLPNVVISAISLPNIRKPDGYDYYAFSKDQYNIDV 1380

Qy 1343 PSRTARAJTTRSGQTLSKVVWNCP 1366
        ||||||||||||||||||||||||
Db 1381 PSRTARAJTTRSGQTLSKVVWNCP 1404
```

```
Qy    143 KIKSSKNSAANRELQKKLKVKDNKKNRIKKKFPTPKPPVVDEAGSGLDNGDFKVTTPDTST 202
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 KIKSSKNSAANRELQKLKVKDNKKNRTKKKFPTPKPPVVDEAGSGLDNGDFKVTTPDTST 240

Qy    203 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLIVNKETTVEIKETTTTNKQTSTDG 262
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLIVNKETTVEIKETTTTNKQTSTDG 300

Qy    263 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS 322
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS 360

Qy    323 TTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKE 382
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEP 382

Qy    383 APTTTKSAPTTPKEPSPTTTKEPAPTTTKEPAPTTPKKPAPTTPKEPTPTTPKEPAPTTPK 442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 APTTTKSAPTTPKEPAPTTPKEPAPTTPKKLTPTTPEKLAPTTPKEPAPTTPKEPAPTTPK 480

Qy    443 EPAPTTAPKKPAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTKSAPTTTKEPAPTT 502
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 EPAPTTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKEPAPTTPKSAPTTTKEPAPT 540

Qy    503 TTTKSAPTTPKEPSPTTTKEPAPTTKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKKP 562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTTKSAPTTPKEPSPTTTKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKEP 600

Qy    563 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKPAPTTPEELAPTTPEEPTPTT 622
          |||||||||||||||||||||||||||||||||||||||||||||
Db    601 APTTPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPKEPAPTTTKEPAPTTPK 660

Qy    623 PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTTTTKGTAPTTLKEP 682
          |||||||||||||||||||||||||||||||||||||||||||||
Db    661 PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTTKGTAPTTLKEP 720

Qy    683 APTTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTTPKGTAPTTPKEPAPTTPKEPAPTTPKG 742
          ||||||||||||||||||||||||||||||||||||||||||||
Db    721 APTTPKKPAPKELAPTTPKEPTSTTCDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKG 780

Qy    743 TAPTTLKEPAPTTPKKPAPKELAPTTKGPTSTTTSDKPAPTTPKETAPTTPKEPAPTTPK 802
          ||||||||||||||||||||||||||||||||||||||
Db    781 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTTSDKPAPTTPKETAPTTPKEPAPTTPK 840

Qy    803 KPAPTTPEPTPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT 862
          |||||||||||||||||||||||||||||
Db    841 KPAPTTPEPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT 900

Qy    863 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 922
          ||||||||||||||||||||||||||||||||
Db    901 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 960

Qy    923 TSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAINSKAATTPK 982
          |||||||||||||||||||||||||||||||
Db    961 TSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAINSKAATTPK 1020

Qy    983 PQKPTKAPKKPTSTKKPKTMPRVRKPKTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1042
          ||||||||||||||||||||||||||||||||||
Db   1021 PQKPTKAPKKPTSTKKPKTMPRVRKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080

Qy   1043 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPVHFMPEVTPDMDYLPRVPNQGTIINPMLS 1102
          |||||||||||||||||||||||||||||||||||||||||
Db   1081 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPVHFMPEVTPDMDYLPRVPNQGIIINPMLS 1140

Qy   1103 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPSPARRITEVWGIPSPIDVFT 1162
          |||||||||||||||||||||||||||||||||||||||||
Db   1141 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPSPARRITEVWGIPSPIDVFT 1200

Qy   1163 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVY 1222
          |||||||||||||||||||||||||||||||||||||||
Db   1201 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVY 1260

Qy   1223 FPKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1282
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 FPKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1320

Qy   1283 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV 1342
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV 1380

Qy   1343 PSRTARAITTRSGQTLSKVWVNCP 1366
          ||||||||||||||||||||||||
Db   1381 PSRTARAITTRSGQTLSKVWVNCP 1404
```

```
RESULT  3

ID        Q9JM99
AC        Q9JM99;
DT        01-OCT-2000 (TrEMBLrel. 15, Created)
DT        01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT        01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE        MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE        PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN        PRG4.
OS        Mus musculus (Mouse).
OC        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX        NCBI_TaxID=10090;
RN        [1]
RP        SEQUENCE FROM N.A.
RA        Ikegawa S., Nakamura Y.;
RT        "a novel mouse gene highly homologous to a human gene encoding
RT        megakaryocyte stimulating factor precursor and cartilage superficial
RT        zone protein.";
RL        Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR        EMBL; AB034730; BAA92310.1; -.
DR        MGD; MGI:1891344; Prg4.
DR        InterPro; IPR000585; Hemopexin.
DR        InterPro; IPR002965; P_rich_extensn.
DR        InterPro; IPR001212; Somatomedin_B.
Pfam;     PF01033; Somatomedin_B; 2.
DR        PRINTS; PR01217; PRICHEXTENSN.
PRINTS;   PR00585; HEMOPEXIN.
DR        SMART; SM00120; HX; 2.
DR        SMART; SM00201; SO; 2.
PROSITE;  PS00024; HEMOPEXIN; UNKNOWN_1.
PROSITE;  PS00524; SOMATOMEDIN_B; 2.
SQ        SEQUENCE  1054 AA;  115991 MW;  4FC64BFA42283235 CRC64;
```

```
Query Match     51.5%;  Score 3759;  DB 11;  Length 1054;
Best Local Similarity  55.0%;  Pred. No. 8.6e-237;
Matches  779;  Conservative  56;  Mismatches 169;  Indels 412;  Gaps 26;
```

```
Qy      1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60
          |  |  |  |  | :|||||||||||||||||||||||||||||||||||||||||||
Db      1 MGWKILPVCLSLLLPVVLIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60

Qy     61 KRVCT---------------------------------AAEVHNPFSPPSSKKAP 82
          ||:|
Db     61 KRVCSPELSCKGRCFESFARGRECDCDSQCKQYGKCCADYDSFCEEVHNSTS-PSSKTAP 119

Qy     83 PPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSTIW 142
          :||  |||||||||||  | : |||||:|::||||||||::|||:|:| |||||||
Db    120 TPAGASDTIKSTTKRSPKSPI-TRTIKVVESEELTEEHSDSENQE--SSSSSSSSSSTIR 176

Qy    143 KIKSSKNSAANRELQKKLKVKDNKKNRIKKKFPTPKPPVVDEAGSGLDNGDFKVT--TPDT 200
          ||||||| ||||:|||||||||||||||:||||:|||||||||||||||:|:|  :| |
Db    177 KIKSSKNS-ANRELQKNPNVKDNKKNTPKKPNPEPPAVDEAGSGLDNGEFKLTPPPPDP 235

Qy    201 STTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLIVNKETTVEIKETTTTNKQSSA 260
          ||| |:::||||||||||:|||| :|| :|:|||::|| |||||||||||||:|:|||
Db    236 PTTPHSKVAISPKTTAAKPVTPKPSLAPNSETSKEASLASNKETTVETKETTATNKQSSA 295

Qy    261 DGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEP 320
          |||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKG------TPDT 320
```

```
              |: |||| |||| ||:                                                                    ||
Qy    296   -SKKKTISVKEIRSAEKTSDKDV------------------------------------EP    319
Db    320   TSTTPK------------NSAPTTKKPV-TTTKESKFLP--------                   357
              |||| |    |:|| ||:      |  |||| |
Qy    321   ASTTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTKEPAPTTK                      380
Db    358   EPPTTKPEPPTTPKEPPTTPKEPEPTTPKEPEPTTPKEPEPTT                      417
              | |||  :|: || |||| ||:| || |||| | :| |||| |
Qy    381   EPAPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT                     440
Db    418   PKEPEPGTTPKEPEPTTPTTKEPEPTTTKREPEPTTPKEP----                     461
              :| ||  || |||| ||:| ||| :| ||||| ||
Qy    441   PKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKSAPTTTKEPA                     500
Db    462   -----EPTTPKEPEPTTLKEPEPTT------TPKEPEPTTPKEP                     512
              |||| ||:| |:|||| || :| ||||
Qy    501   PTTTKSAPTTPKEPSPTTTPKEPAPTTPKEPAPTTPKEPAPTTK                     560
Db    513   ---------KPAPTTAPKEPAPTTPKELAPTTPKETAPTTPKGTAPTTLK               512
              ||||  || |||| ||:| |||| |||| ||:| |||| ||
Qy    561   KPAPTTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPKEPAPTTPEEPTP            620
Db    513   -------------EPTTPKEPEPTTPKETAPTTPKGTAPTTLK                      512
              |||| || |||| ||:|
Qy    621   TTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLK             680
Db    540   -------------EPTTPKEPVPTTPKEPEPTTP                               570
              || |||| ||
Qy    681   EPAPTTPKKPAPKELAPTTPKEPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKEPEPTTP     740
Db    571   -------------KEPEPTTPKEPEP                                       609
              :| |||| || ::|
Qy    741   KGTAPTTLKEPAPTTPKKPAPKELAPTTIKGPTSTTTSDKPAPTTPKETAPTTPKEPAPTT    800
Db    610   PKKPEPTT-------------------TTRKEPEPTTPKEP                        617
              |||| |||
Qy    801   PKKPAPTTPTPEPPPPTTSEVSTPTTTIHKSPDESTPELSAEPTPKALENSPKEPGV        860
Db    618   PTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTTESKITATTT    662
              |||| |||
Qy    861   PTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTTESKITATTT    920
Db    623   --------TLTKATTLAPKVTAPAE----EIQNKPEETTPASEDSDDSKTTL             719
              |||
Qy    921   QVTSTTTQQDTTPFKITTTLKTTTTLAPKVTTTKKTTITTTEIMNKPEETTAKPKDRATNSKAT  979
Db    663   KPQKPTKAPKKPTKAPKKPTSTKKPKT--PKTRKPKTTPAPLKTTSATPELNTTP--L         778
              |||
Qy    980   ----------TPKP-QKPTKAPKKPTSTKKPKTMPVRVRKPKTTPTPRKMTSTMPELNPTSRIA  1031
Db    663   KPQKPTKAPKKPTKAPKKPTSTKKPKT--PKTRKPKTTPAPLKTTSATPELNTTP--L         719
              |||
Qy   1032   EAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVP      1091
Db    720   EVMLPTTTTIPKQTPNPEIAEVNPEHEDADGGEGEKP-LIPGPVLFPTAIPGTDLLAGRL      778
              |||
Qy   1092   NQGIIINPMLSDETNICNGKPVDGLTTLLRNGTLVAFRGHYEFWMLSPFSPPSSPARRITEVW   1151
Db    779   NRGININPMPSDETNLCNGKPVDGLTTLLRNGTLVAFRGHYEFWMLNPFRPPSPPRRITEVW    838
              |||
Qy   1152   GIPSPIDTVFTRCNCEGKTFFFKFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALST   1211
Db    839   GIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDVVDPGYPKQTVKGFGGLTGKIVAALSI     898
              |||
Qy   1212   AKYKNWPESSVYFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIG      1271
Db    899   AKYKDRPESVYFFKRGGNIQQTYKQEPMKKCTGRRPAINYSVYGEAAQVRRRFERAVG       958
              |||
Qy   1272   PSQTHTIRIQYS-PARLAYQDKGVLHNEVKVSILMRGLPNVVTSAISLPNIRKPDGIYDY     1330
Db    959   PFQTHTIRIHYSVPMRVSYQDKGFLHNEVKVSVSTMWRGLPNVVTSAITLPNIRKPDGIYDY    1018
              |||
Qy   1331   AFSKDQYYNIDVPSRTARAIITTRSGQTLSKVWVNCP                            1366
```

```
Query Match              23.5%;     Score 1713.5;  DB 6;  Length 401;
Best Local Similarity   80.2%;     Pred. No. 4.8e-104;
Matches 320;   Conservative 24;  Mismatches 52;  Indels 3;  Gaps 3;

Qy    969   PKDRATNSKATTTPKPQKPTKAPKKPTSTKKPKTMPVRVRKPKTTPTPRK-MTSTMPELNPT    1027
              ||  |||||  | ||||||||||  |||||||| |          | |||||||||||
Db      5   PKGRATNSQVTTPKPQKPTKAPKPKPTSTKKPRT--PRVRKPKTTPTPPPKTTTSAMPEPTPT    63
Qy   1028   SRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYL     1087
              :  |||||||||||: ||||| ::|||| || :|:||| ||| ::|| |
Db     64   S-LPEAMLQTTTRPTPTPNSEIIDVNSENEDGDAAEGEKPHMIFRPPVLTPIVILPGTEII     122
Qy   1088   PRVPNQGIIINPMLSDETNICNGKPVDGLTTLLRNGTLVAFRGHYEFWMLSPFSPPSSPARRITEVW   1147
              :  ||| |||||| |||||:|||:|||||||||||||||||||||||||| || | | ||||
Db    123   VRGPSQGFGINPMFSDETNLCNGRPVDGLTTLLRNGTLVAFRGHYEFWMLTPFTPPPPRRI     182
Qy   1148   TEVWGIPSPIDTVFTRRCNCEGKTFFFKFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVA    1207
              |||||||||||||| ||||||||||||||| ||||||||||||||||::| |||||| | |||
Db    183   TEVWGIPSPIDTVFIFRCNCEGKTFFFKFKGSQYWRFTNDIKDAGYPKLISKGFGGLNGKIVA    242
Qy   1208   ALSTAKYKNWPESSVYFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRRFE     1267
              |||:|||| |||| ||||||||||| ||||||||||| ||||:|| |||||||| ||  ||
Db    243   ALSIAQYKSRPESVYFKRGGSVQQYIYKQEPTQKCTGRRPAINYSVYGEFAQVRRRRFE     302
Qy   1268   RAIGPSQTHTIRIQYSPARLAYQDKGVLHNEVKVSILMRGLPNVVTSAISLPNIRKPDGY     1327
              ||||||| ||||| | |: | ||||:|||||||||  |||||||||||||:||||||||||
Db    303   RAIGPSQVHTIRIHYTPVRVPVQODKGFLHNEVKVSTLMRGLPNVVTSAITLPNIRKPDGY     362
Qy   1328   DYYAFSKDQYYNIDVPSRTARAIITTRSGQTLSKVWVNCP     1366
              | || |||||||||||||||||||||||||||||: ||||
Db    363   DIYALSKDQYYNIDVPSRTARAIITTRSGQTLSNTWVNCP     401
```

```
Db   DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
     DE   Y51B11A.1 PROTEIN.
     GN   Y51B11A.1.
     OS   Caenorhabditis elegans.
     OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
     OC   Rhabditidae; Peloderinae; Caenorhabditis.
     OX   NCBI_TaxID=6239;
     RN   [1]
     RP   SEQUENCE FROM N.A.
     RC   STRAIN=BRISTOL N2;
     RX   MEDLINE=99069613; PubMed=9851916;
     RA   None;
     RT   "Genome sequence of the nematode C. elegans: a platform for
     RT   investigating biology. The C. elegans Sequencing Consortium.";
     RL   Science 282:2012-2018(1998).
     RN   [2]
     RP   SEQUENCE FROM N.A.
     RC   STRAIN=BRISTOL N2;
     RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
     RA   Cotton M.;
     RN   [3]
     RP   SEQUENCE FROM N.A.
     RC   STRAIN=BRISTOL N2;
     RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
     RA   Waterston R.;
     DR   EMBL; AC006797; AAF60743.1; -.
     DR   InterPro; IPR002965; P_rich_extensn.
     DR   PRINTS; PR01217; PRICHEXTENSN.
     SQ   SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match       Best Local Similarity   29.6%;
Matches 332; Conservative 131; Mismatches 481; Indels 177; Gaps 44;
Score 982.5;  DB 5;  Length 1079;
Pred. No. 5.3e-56;
```

```
Qy   157 QKKLKLVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKV------------TTPDITSTTQ-HNKV 208
             |:|| |||| |       |||  |                    |:|: | ||:|||
Db    27 QKELEIADCRANQTPHTMLP-------------------STLTSVDMEFPSTLVLSSTPTSSSTPIKETT 77

Qy   209 STSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETT-VETKETTTTNKQTSTDGKEKTT 267
          :|:|| || |  | ||                | |||: || | | || |
Db    78 TTAPETTSTEP------------PSSSTTPVQTTTTTAPETTSTSTEPPSSSTTPVQTT-- 124

Qy   268 SAKETQSIEKTSAKDLAPTSKVLAKPTPKRAETT-TKGP-ALTTPKEPTPTTPKEPAPTTP 325
             ||| ||    |||  |  ||    |||||: |||  |||:|||||||||||||||
Db   125 TAPETTSTEPPSSS---------TSPVQTTTTTAPETTSTEAPSSSTTPVQTTTTTAPETTSTEP 180

Qy   326 KEPPTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 385
               ||| |||||||||||                |||||||||||||||||||||||
Db   181 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPE---TTSTEP-PSSSTTPVQT 236

Qy   386 TTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKEPTPTTPKEPAPTTPKEPA 445
          ||| |||||||||||||||||||||||||||| |||| |||||||||||
Db   237 TTTTAPETTSTESPSSSTTPVQTTTTTAPETTSTEP-PSSSTTPVQTTTTTAPE 295

Qy   446 PTAPKKPAPTTPKEPAPTTPKEP-----------------APTTTKEPSPTTPKEPAPTTKS 491
             | ||||||||||||                        |||:| ||| ||
Db   296 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTT 355

Qy   492 APTTTK--EPAPTTTKSAPTTPKEPSPTTTKEP-APTTPKEPAPTTPKKPAPKELAPTTTKGPPTSTTSDKP- 548
          |||:|   | | |||||||||||  |||||||:  |||||||||||      :||||||||||||||||||
Db   356 APETTREPPSSSTTPVQNTTTTTAPETTSTEPPSSSTTPVQTTTTT---APETTSTEPP 411

Qy   549 TTPKEPAPTTTKKPAPTAPKEP--APTTPKETAPTTTPKKLTPTTPEKLAPTTPEKPAPTT 606
             :|||             ||||   || |||||||||
Db   412 SSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEP----APETTSTEPP 467

Qy   607 PEELAPTTPEEPPTTPEEPAPTTPKAAAAPNTPKEP--APTTPKEPAPTTPKEPAPTTPK 664
             |||||                        |||||  |||||||||||||||||||
Db   468 TTTTAPETTSTEP-PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPE-TTSTES 525

Qy   665 ETAPTTPKGTAPTTLKEPAPTTKKPAPKELAPTTTKEPTTSTTSDKP-APTTPKGTAPTT 723
             |||||||| ||||||||||||                     ||||||||||
Db   526 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTT 585

Qy   724 PKEPAPTTPKEP--APTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPPTSTTSDKP- 780
               |:|||||     |||||||||||||||||                  | |||||||
Db   586 APE---TTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPS 642

Qy   781 APTTPKETAPTTPKEPAPTTPKK-----PAPTTPETPPPTTS------EVSTPTTTKE 827
             |||||| ||||||||||||          | ||                ||||||||
Db   643 SSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAP 702

Qy   828 PTTIHKSPDESTPELSAEPT--PKALENSPKEPGVPTKTPAATKPEMTTTAKDKTTERD 885
              |||||  | |:|||   |     ||  ||||  |  || |||   ||||  | ||||
Db   703 ETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTSTPVQTTTTT 762

Qy   886 LRTT----PETTTAAPKMTKETATTTEKTTESKIATTTQVTVSTTTTQDTTPFKITTTLKTT 941
             |:||       ||||       || |||        |||| || ||
Db   763 VQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSNTPVQTTTTTAP 820

Qy   942 TLAPKVTTTKKTITTTTEIMNKPEETA-KPKDRATTNSKATPTPKPQKPTKAPKKPTSTKKPK 1000
             |||| ||| |||    |        ||| ||| | || |  |||||
Db   821 STEPPSSSNTPVQTTT--TTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPV 878

Qy  1001 TMPRVRPKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAG 1060
             |     | ||  |  ||||       |  ||  |||||
Db   879 QTTTTTAPETTSTEPPSSSTTPVQTTTTTAPET---TSTEP----PSSSTTPVQTTTTTAP 932

Qy  1061 GAEG-ETPHMLLRPHVFMPEVTP---DMDYLPRV----------PNQGIIINPMLSDETN 1106
             | |  |                   |  |               |   | | |  ||
Db   933 ETTSTEPPSSSTTPVQTTTTTVPCDCSLSYIDRVVYPFTTEWWENKRDIIIQSYDSPRRT 992

Qy  1107 ---ICNGKPVDGLTTLRNGTLVAFRGHYFWMLS--------PFSPPS 1142
             |                                        |
Db   993 AFSFVSTFDIGTCTRAILTCIYTSEGGISNL-NATL-----FIGLSDGGSIDLPF----- 1040

Qy  1143 PARRITEVWGIPSRIDTVFTRCNCEGKTFFFKDSQYWRFTN 1183
             |                         | |   |
Db  1041 ------------YNVPIGLE-IWPEINCEGNK-------WSYNN 1064
```

RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

```
RESULT   6
ID   Q9I7S1            PRELIMINARY;      PRT;   1049 AA.
AC   Q9I7S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
```

RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003495; AAG22353.1; -.
DR FlyBase; FBgn0030561; CG5228.
SQ SEQUENCE 1049 AA; 107278 MW; 954D6292E7619671 CRC64;

Query Match 13.0%; Score 946.5; DB 5; Length 1049;
Best Local Similarity 34.4%; Pred. No. 1.1e-53;
Matches 376; Conservative 69; Mismatches 480; Indels 169; Gaps 56;

```
Qy      68  EVHNPTSPPSSSKAPPSGASQTIKSTTKRSPKPPNKKKVIESEEITEEHSVSENQE 127
                :|  ||:  | :                            :|| |||||
Db      33  EIGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPT----------EGTTAKPTTLKPTE  86

Qy     128  SSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTPKPPVVDEAGSG 187
                                          :|:|  |  ||:||      :  || ||
Db      87  GTTAKPTT--------LKPTEGTSA----KPTTLKPTE--GTTAKPTTLKPT---EGTS 128

Qy     188  LDNGDFKVTTPDTS--TTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLIVNKETT 245
                 :|  || ||  |  |:| ||  |||||||||||||||||||||
Db     129  AKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKP---TTLKPTEGTSAKPTT 185

Qy     246  VETKETTTTNKQT--STDGKEKTTSAKET--OSIEKTSAK--DLAPTSKVLAKPT---PK 296
                |  | | :|  |   | |  |  || |  ||| | |  :|            |  ||
Db     186  LKPTEGTSAKPTTLKPTEG------TSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPT 241

Qy     297  AETTTKGPALTTPKPEEPTTTPKEPASTTPKEPT---PTTIKSAPTTPKEPA---PTTTKS 350
                |  ||  | || |||| | ||| |||| | | | |  ||:||| | |  ||:|   :||||||
Db     242  EGTTAK-----PTTLKPTEGTTAKPTTLNPTEGTSAKPTTLKPTEGTTAKPTTLNPTEGTS 297

Qy     351  APTTPKEPAPTTPKEPA----PTTPKEPAPTTTPKEPAPTTPKEPAPTTKKPAP 407
                |  ||  | :||   | ||: || |  ||| |  | | ||| | |:|||   :|| |
Db     298  AKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLK--PTEGTSAKP 352

Qy     408  TTPKEPAPTTPKEPT--PTTPKEPAPTTPKEPAPTTPKEPA---PIAPKKPAPTTPKEPA 462
                 ||||||||| |  | ||:|  |  :||| | || | ||| |   | | |||| ||:||:
Db     353  TTLKPTDGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG 412

Qy     463  TTPPKEPAPTTTKEPSPTTTKEPAPTTTKSAPTTPKEPA---PTTTKSAPTTPKEPSPTTT 519
                | | | ||| | ||   |||| ||| ||:|| | |  ||   ||:||| | |||   |||
Db     413  TTAK---PTTLKPTEGTTAK---PTTLKPTEGTTAKP 466

Qy     520  KEPAPTTPKE---PAPTTPKKPEPAPTTPKEPAPTTKKPAPTAPKE---PAPT 573
                ||||||||| |   | || | | | ||  ||| | ||:||| |  |   |||
Db     467  AKPTTLKPTEGTTAKPTTLK-PTEGTTAKPTTLKPTEGTSAKPTT 525

Qy     574  TPKETAPTT--PKKLTPTTPEKLAPTT--PEKPAPTTPEELAPTT--TPEEPTPTTPEEPA 627
                ||||||||   ||:|   || || ||   ||:|  ||  || ||   |||| |||||||
Db     526  TLKPTKGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGT 585

Qy     628  PTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTPLKPTEKEPAPTTP 687
                ||||:| | | ||||||| |||||||| | :||| |     ||| | ||| | ||| |||||
Db     586  TAKPTTLKPTEGTSAQPTTLK---PTEGTSAQPTTLKPTEGTTAK---PTTLK---PTEG 636

Qy     688  KKPAPKELAPTT-----------TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEATPP 740
                ||||||||||           |:||| | ||: | | ||| | ||| ||||||:| | |||
Db     637  TSAQPTTLKPTEGTTAKPTTLKPTEGTSAK---PTTLKPTEGTTAK---PTTLKPTDGTTA 691

Qy     741  KGTA-------------PTTLKEPAPTTPK----KPAPKELAPTTLKPGPTSTTSDKPAPTTPKT 786
                |||             | ||  || ||||    ||||||| || ||| ||| |   | |||
Db     692  KPTTLNPTEGTSAKPTTLKPTEGTTAK---PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPT 751

Qy     787  ETAPTTPKEPAPTT--PKKPAPTTPEPTEPPPTTSEVSTPPTTTK----EPTTHKSPDE 837
                || ||| |||||    |  | || |    |||| ||||   |||       ||||||:||
Db     752  EGTRTTAK---PTTLKPTEGTSAKPTTLKPTEGTTAK------EPTTHKSPDE 806

Qy     838  STPELSAEPTP-KALENSPKEPGV--PTTKT---PAATKPEMTTTAKDKTTERDLRTTPE 891
                ||| | || |  ||||| || |    |||||   |||| | | |:|||| ||||||||||
Db     807  RT---SAQPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTSAKPTTLKPTEGTTAKPTT----LKPT-E 858

Qy     892  TTTAAPKMTKETATTTEKITTESKITATTTQVTSTTTQDTTPKITTLKTTTLAPKVTTTK 951
                ||| | |  |:||||:| |||| |||||  ||  | | |||||| |||| ||||  |||
Db     859  GTSAKPTTLKPTEGTTAK----TLKPTEGTTAKPTTLKPTDGTTA 914

Qy     952  K--TITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPK---KPT--STRKPKIMPR 1004
                    |:| |:||:|  ||| :|| | ||:|| |||| | |    ||   |:|||||||:
Db     915  KPTTLKPTEGTSAKPTTLKPTE-GTTAKPTTLKPTEGTSAKPTTLKP 973

Qy    1005  VRKPKTTPT---PRKMTSTMP-ELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAG 1060
                :|||::|||   | |||  |  ||||| || | ||| | || || || || ||| ||
Db     974  TEGPSAKPTTLKPTERTSAQPTTLKPTEGTTAK---PTTLKPTEGTTAKPTTLKPTKGTSG 1031

Qy    1061  GAEG-ETPHMLLRP 1073
                  |  ||| |||||
Db    1032  QANNPETKRKERRP 1045
```

RESULT 7

ID   O76894
AC   O76894;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:5G7.1 PROTEIN.
GN   EG:5G7.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Cadieu E., Dreno S., Lelaure V., Mottier S., Galibert F.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE003421; AAF45644.1; -.
DR EMBL; AL031028; CAA19845.2; -.
DR FlyBase; FBgn0025390; EG:56G7.1.
DR InterPro; IPR002557; Chitin_binding.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF01607; Chitin_bind_2; 2.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00494; ChtBD2; 2.
SQ SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match   12.2%; Score 891.5; DB 5; Length 1795;
Best Local Similarity   27.0%; Pred. No. 7.5e-50;
Matches 317; Conservative 154; Mismatches 510; Indels 195; Gaps 40;

Qy   36 EGYSRDAT-CNCDYNCQ-------HYMEC-----CP-----DFKR-VC--------TAAEV 69
                          | | | | | | |
Db 224 EGTERSPTDCALYTCRLQESGTIYLQTRFKCPGSNSFDLERKLCRPRSEVDCFDFVPGPV 283

Qy   70 HNPTSP------PSSKKAPP-------PSGASQ---TIKSTTKRSPKPPNKKKTKKVIESEE 115
           || ||| | :| | ||| || | || | ||
Db 284 QVPYAPQPYYPYPYPAAPPLYEEDDYDTGAREQQPALKSEKLQVAAEGFEKPSLNVVVLQT 343

Qy 116 ITEEHSVSENQESSSSSSSSSSSSSSSSTIWKIKSSKNSAANRELQK------KLKVKDN---- 165
                : || || | | | | ||| ||
Db 344 TTLEPSTAYHKYPAYPSYPSYEYSS--HHRGKERAAENLELEKEGVPRKLKLSENIVIQP 401

Qy 166 ------------KKNRTKKKPPTPKPPVVDEAGSGLDNGDFKYTT 197
                       | | | | | | | | | | |
Db 402 ETPATAATTREPLNDINKYQKRYTYGTDKNDVTEAPEIKSPL----KGLHLSENIVIL 456

Qy 198 PDTSTTQHNKVSTSPKITTAKPINPRSLPPNSDT-----------SKETSLTVNKETTV 246
        | | : | || || | || | || | ||| |||
Db 457 PETTTT-TTTTTKPVVLTCPIISPPDTTPKPSTTTAVTKSTPKISSTEQHSTTTAKTTT 514

Qy 247 ETKETTTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPAL 306
        | | |:| | | ||| | | | | : | | | || | | || | || |
Db 515 TKRPTTVTEKTSSATEKPRTTVVTTTTQKRSTTTHNTSPDTTKITTIRSTTLSPKTTTPST 574

Qy 307 TTTPKEPPTPTPKEPASTTPKPEPTTTIKSAPTTPKEPAPTTTK-E 365
        || | || | || || | | | || ||| ||||| | |
Db 575 TTPSTTTPST----TTPSTTTPSTTTTVKVSTHRPRTTSQKTT 623

Qy 366 PAPTTPKEPAPTTTKKSAPTTPKPRAPTTTKEPAPTTTKEPRPTTT 425
        |||| | || | | || | | ||| ||||||| ||| || |
Db 624 TASTTTKK------TTTSPKTTKTDIPTSTTSKLSTTTQKTTTTTHKFTAATTSTEKPKTT 679

Qy 426 PKEPAPTTPKEPAPTTTKEPAPTTTKKPAPTTTKEPAPTTTKEPAPTTTKEPSPTTPKEPA 485
        | | | | || | || | | | | || | | || | | || || | |
Db 680 TEKTSTVSTTVSTTPKKKSTESSSPKPTSSSTGKPTTTTPKPSTRTTTPTTTKVTTTTQITTTTPLRSS 739

Qy 486 PTTTKSAPTTPKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTTKK--------- 536
        || | | || | || | | || | | | || | | |||| ||
Db 740 TETTTSTQPPTTTTTPQPTTTTTLTVTPKTSTTTTTTEKPITTSSSPKPTTTTSTTAPNTT 799

Qy 537 KPAPTTPKEPAP----------TTPKEPAPTTTKKPAPTAPKEPAPTTPK-ETAPTTP 583
        || || | || ||| || | | || | ||| | | | || |
Db 800 KVAITTQKETTPTQSTTTETTFTTPKTTTNNEPTTSTEKPIISTTPKTSTVASST 859

Qy 584 KKLTPTTPEKLAPTTPEKP------------APTTPEELAPTTTPEEPTPTTEEPAPTTPKAAA 635
        || | | || | | ||| | | || | || || |||
Db 860 EKTTISSPKPTTEKSTENPTTNSVKTSALTSSTQRATSTTSSEPTKTT-QNTTTTPKPTT 918

Qy 636 PNTPKEPAPTTPKEPAPT---TPKEEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPK 693
        | ||||| || || | | || ||| | |||| ||| | ||| | | | |
Db 919 LKTSTQEATTSTQKVSTVTTIITTKKATESSPLTTLSEEPNTTPKPLRTTTPTTTSVTATT 978

Qy 694 ELAPTTKEPT--STTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEP 751
        || || | | | | ||| | ||| || ||||| || | || || ||| || | |
Db 979 RITTTTISESSTETTTSTQRPKSTTPTSTTTRTTTPKVTTTVIVSTQNPTTTTSKTSTVTI--- 1035

Qy 752 APTTPKKPAPKELAP-TTTKGPTSTTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPE 810
        ||| | || | || ||| ||| || | || | ||| || | ||| | || |||
Db 1036 --TTP-NPSFSTQRPTTTIRQPISITAASTTSIGTTRIPTTTNPQNSTSSSTDLTTVTRPC 1092

Qy 811 TPPPTTSEVSTPTT-TKE------------PTTIHKSPDESIPELSAEPTPKA 850
        | | | | | | || | | | | | | ||| |
Db 1093 PDPDSTTSDKNTNTACTTQELCQVNLLELQSPQKQEQFTHTRTHTALTGSRNTLGGOEVPDY 1152

Qy 851 LENSPKEPGVPTTKTPAAATKPEMTTTAKDKMTTER---DLRTTPETTTAAPKMTKETATTT 907
        | || || | | | | | | | | | | | | | | || || | || ||
Db 1153 MDDAPSSAEAESGQATTAKAPTMSTLAAAHLLQKLFHIISTTPPSREHAP--TQRPSSQP 1210

Qy 908 EKTTESKITATTTQVTSTTTQODTTPFKITTLKTT---TLAPKVTTTKKITTTTEIMNKP 963
        ||| | || | | || | | || |||| ||| || | ||||| | | || |
Db 1211 SSSQQRSR-GVTIAQMARHNLATSKPFIAHSLRLSIIQQLLASTQKRSIPPKLVTHNTTKEP 1269

Qy 964 EETAKPKDRATNSKATTPKPQKPTKA-PKKPTSTKSPKPKTMPRVRKPKTTPTTPRKMTSTMP 1022
        ||| | | | || || | | | ||| | || | | | || | || || |||
Db 1270 ED-SEYVDSETSEQYTDEDNEVLDKTQPRAMSSTTVAAVLPAVPSTTTEREPQK-TSSSP 1327

Qy 1023 ELNPESRIAEAMLQTTTRENQTPNSKLVEVNPKSED 1058
        || | | | ||| | | | | | | | | ||
Db 1328 --SPT----KATSSTTTQPIETTTGDLEVDSGSGSSD 1357

RESULT 8
ID   O96449
AC   O96449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RACE 1-11;
RA   Goenhardt B.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061105; AAC72308.1; -.
SQ   SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

O96449   PRELIMINARY;   PRT;   1489 AA.

```
Query Match    11.8%;  Score 863.5;  DB 10;  Length 1489;
Best Local Similarity   31.7%;  Pred. No. 4.1e-48;
Matches  368;  Conservative  55;  Mismatches  509;  Indels  229;  Gaps  51;

Qy    107 TKKVIESEITEEHSVSENQESSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKILKVKDNK 166
          ::: :   :    ::    ::    ::    ::   ::  :::
Db    282 TPTVGYSTEETEGQHVTGGYEPSDETEAPTEGTTV--VPREETTAAPSE------DTTYAP 334

Qy    167 KNRRKKPPPKPPVVDEAGSGLDNGDFKVTTPDTSTT-----QHNKVSTPKITTAKPI 220
          ::      :::                          :::
Db    335 REVTPYAPTEEKPVDVEETTYVTEESTYAPTKSEINAPTERMHYAHIEKPCDTTEVTMYAPT 394

Qy    221 NPRPSLP-------------SKETSLIVNKETTVETKETTT----TNKQT 258
          :::                 ::::::
Db    395 EETTYAPTEETTYAPTEEPYEPTEENTTYPTEENTTYAPTEEPTTPK 454

Qy    259 STDGKEKTYSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPK 318
          :::  :::::
Db    455 YAPTEETTYAPTEETTYEPTEENTTYAPTEENTTYAPTEEPTTPK---EETTYASTEETTYAPTEETTY 511

Qy    319 EPASTTPKEPTPTTIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT----TTKEPAPTTPKEPA 375
          :::    :::                    :::
Db    512 APAEETPYEPTEET-TYAPTEEETTYAPTEETTYAPAEEETPYEPT 570

Qy    376 PTTTKEPAPTTTKSAPT----------TPKEPAPTTPKPAPTTPKEPAPTTPKEPTPTT 425
          :::           ::
Db    571 EETTYAPTEETT-YAPTEETMYAPIEETTYAPTEETTYAPAEEETPYEPTEETTYAPTEET 629

Qy    426 PKEP---------AP--TTKEPAPTTPKEP------APTAPKKPAPTTPKEPAP 462
          :::
Db    630 TYAPTEETTYASTEETTYAPTEETTYAPAEEETPYEPTEETTYAPTEETTYAP 689

Qy    463 TTPKEPAPT----TTKEPSPTTPKEP-----APT-TTKSAPT------TTKEP--- 499
          :::
Db    690 TEETTYAPTEETTYAPEEETPYEPTEETTYAPTEETTYAPAEETPYAPTEETTYGPTEETT 749

Qy    500 --APT--TTKSAPT---TPKEPSPTTPKEP-----APTTPKEPAPTTPKKPAPTTPKEPAP 548
          :::
Db    750 TYAPTEATTYAPTEEETPYAPTEETTYEPTGEETTYAPTEETTYAPTEEETTYAP 809

Qy    549 T--TPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTT 606
          :::
Db    810 TEETPYEPTEETTYAPTEEEPTPYEPTEET---TYTPTEETTYAPTEETTYAPTEKTTYAPTE 867

Qy    607 PEELAPT--TPEEPTPTTPEEP-----APT------TPKAAAPNTPKEPAPTTPK 648
          :::
Db    868 EETTYAPTEETPYEPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEEPTY 927

Qy    649 EP------APTTPKEPAPTTPKETAPTTPKGTAPT---TLKEPAPTTPKKPAPK-ELAPT- 698
          :::
Db    928 EPTEETTYAPTEETTYAPKGTAPTTPKEPAPTTPKEP-----APTTPKGTAPT----- 987

Qy    699 -TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP------APTTPKTAPT 746
          :::
Db    988 EETTYAPTEEETMYAPIEETTYAPAEEETPYEPTEETTYAPTEETTY 1047

Qy    747 -------TLKEPAPTTPKKP-----LAPT--TTKGPTSTTSDKPAP 782
          :::
Db   1048 ASTEETTYAPTEETTYAPAEEETPYEPTEETTYAPTEETTYAPTE 1107

Qy    783 TTPKETAPTTPKEPAPTTPTPEPTPPPTTSEVSTPTTTKEPT--TIHKSPDESTP 840
          :::
Db   1108 EETTYAPAEETPYEPTEETTYAPTEETTYAPIEETTYGPTEETTY--APTEATT 1166

Qy    841 ELSAEPPPKA-LENSPKEPG-----VPTTKIPAATKPEMTTTAKDKTTERDLRTTP---- 890
          :::
Db   1167 YAPTEETPYAPTEETTYEPTGPTTEGETTYAPTEETTYAPTEEPTYEPT---- 1226

Qy    891 ETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTT--------------- 937
          :::
Db   1227 EETTYAP--------TEETTYEPTEETTYAPTEETTYAPIDEETTYGP 1281

Qy    938 LKTTTLAPKVTTTKKITTTEIMNKPEETAKPKDRAT--NSKATTPKP-QKPTKAPKKPT 994
```

```
Q20007
ID   Q20007       PRELIMINARY;       PRT;   1274 AA.
AC   Q20007;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   COSMID F35A5.
GN   F35A5.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifkin L., Roopra A., Saunders D.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Leimbach D.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U46675; AAB52641.1; -.
SQ   SEQUENCE   1274 AA;  138065 MW;  07F6B0D292C1799F CRC64;
```

```
Db   1282 TEETTYAPTEATTYAPTEETPYAPTEETTYPGETTYAPTEETTYAPTEETTYAPMEET 1341
          :: :    :  :: :     :::   :: :
Qy    995 -------------STKKP-KTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP 1041

Db   1342 PYEPAEEESTSTVUSTEKPCNTEEFTDEPTDEPTDE--PSDEPTDEPTDEPTDLPTDEPSTP 1399
          ::::   ::::    :::     :: :: :        :: ::  :: ::
Qy   1042 --NQTPNSKLVEVNPKSEDAG 1060

Db   1400 CDNQGINGIGVENKVRYNNAG 1420
          ::   :      : :
Qy   1061 ...
```

```
Qy     76 PSSKKAPPSGASQTIKSTTKRSP-------KPP----NKKKTKVIESEITEEHSVSENQES 128
          :::  :::   : ::    :: :          :::    :::
Db    185 PSPKKAAPSKEHDPIVPPTIKNPAKKWKKPPWEDDEVPTEEIKEPEPATRKVPALKKKEP 244

Qy    129 SSSSSSSSSSSTIWKIKSSKNSAANRELQKILKVKDNK-KNRTKK-KP-----TP----- 176
          ::::::::::::
Db    245 STSVKPVSDSPTPKKV---------PVKKEPEVPPPIKNPTKKWKWPPWEDETPVEEVK 294
```

```
RESULT  10
ID   Q9SPM0       PRELIMINARY;      PRT;    1315 AA.
AC   Q9SPM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN.
GN   PEX2.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=POLLEN;
RA   Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA   Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT   "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159297; AAD55980.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR001998; Xylose_isom.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR   SMART; SM00370; LRR; 3.
SQ   SEQUENCE   1315 AA;  134401 MW;  64C97A2A01F0936F CRC64;

Query Match         11.2%;  Score 817;  DB 10;  Length 1315;
Best Local Similarity   23.4%;    Pred. No. 3.9e-45;
Matches  269;  Conservative  109;  Indels  282;  Gaps  37;
```

```
Qy   177  KPPVVDEAGSGLDNGD----FKVITPDTSTTQHNKVSTSPKITTAKPI-NPRPSLPPNSD  231
          ||||  :  :                :  |   :     |  |   |          :
Db   295  EPPVPEKKAPVLKKDPAPAAAKARDPSPSKRAAPKKVEPSSPVVPPTVPKNPVKKPKPPWE  354

Qy   232  TSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLA  291
              :                          ::
Db   355  VDDEPEAEEVKKPSAPEKKTPVLKRKEPEPSSTTPSSDPSPKKAAPAVKPRDSSPKKATPL  414

Qy   292  KPTTPKAE----TTTKGPA-----------------------------------------  326
          | ||  |     | |  |                                      ||
Db   415  QADPKAQEVPPTPVKNPVKKYKPWEVDDEDPEVEVKQPEPAAPAKKTPVLKRKEPAAKDTA  474

Qy   327  EP----TP-TTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPA  367
          ||    || ||| |  |  :||  |   :     |  |   |
Db   475  KPATSKTPEFPEKDPVKRDSSPKKVAAKPDSAQAPA-TPVKNPVKKWRPPWEDDDETPA   533

Qy   368  -----PTTPKEPAPTTTKEPAPT----------TTKSAPTTPKEP-----APTTTPKKPAPTTP  410
               ||||||  | | | |            |||| |||||||
Db   534  DDVSKPTDAKKTPSLAKKDPAPAAPAKESLKPKADTKAPAKPRDSPKKVAPTAPEKKTPVLA  593

Qy   411  KE----PAPTTPKEPTTPTPKEPAPTTTKEPAPTTPK------EPAPTTAPKKPAP----  459
          ||    ||||  |  |   |  |   |   :
Db   594  KKEPAGPADSKTKEPEKSKPRDPSPKKAVPAVPKTEVAPAAVKPEPISKPKDTAPKK     653

Qy   460  PAPTTPKEPAPTTTKEP---SPTTPKEPAPTTTKEPAPTTPK----EPAPTTTKSAPTTPK  514
          || |  |  |  | |       ||||||  | | | |
Db   654  AEPNSPVVP--PTPVKNPVKKWKPPWEDDDAPAKPVSLPEPEKK-TPVLAKKAPIKPDSEA  711

Qy   515  SPTTTKEPAPTTPK---EPAPTTPKKPAP--TTPKEPAPTTPKEPAPTTTKKPAPTA-PKE  569
          || | | ||||  |   || |   | |     |  |  |
Db   712  AADPVSGPSSKDPKLAKKAPVKPRDPSPMKAVPIKPAPKT--EVPPAAVKKPEPVAKSRD  769

Qy   570  PAPTTPKEPAPTTTPKKLTPTTP-----EKLAPTTPKEPAPTTPEELAPTTPEEP-----T  619
          || |  |  |   |  |          |  |   |  |
Db   770  PSPKKAK-AEPNSP--VVPPTPVKNPVKKWKPPWEDDDAPAEPVNVPEPEKKTPVLAKKT  826

Qy   620  PTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTT--PKEPAPTTPKETAPTTPKGTAPTT  678
          |   | || |  |     | ||  |   |  |       |  |  |
Db   827  PVKRDPSPKKAVPAKPSRKTDAPPVSVKKPEPVSKPKEPSPKKAEPNSPVVP-----PTP  882

Qy   679  LKEPAPTTPKKPAPKELAPT-TTKEPTSTTSDKPAPTTTPKGTAPTTPKEPAPTTPKEPAP  737
           ||| |  | | |  | |     | |        |   |  |
Db   883  VKNPVKKW-KPPWEDDDDEPTEEVKKPSE--PEKKTPVLAK-KEPEKPKD-APKVAAKPRD  937

Qy   738  TTPKGTAPTTLKEPAPT-------TPKPAP-------KELAPTTKGPTS-----TTSDK   779
          |  |    |  |  | |         | |         | | |         || |
Db   938  PSPKKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEAPPTPVKNPVKKWKPPWEDDDE  995

Qy   780  P-----APTTPKEP-------APTTPKKPAP----TTPEPTPPPTTSEVSTPTTTKEP  828
                | | | ||         | |   |       ||  | |  |
Db   996  PAEPVSAPEPEKKTPVLAKKAAPAKPRDP---SPKKAAPVAAK-PDDPKIPEV-PPTPVKNP  1050

Qy   829  TTIHKSP----DESTPELSA-EP----TPKALENSPKEPGV------  867
           |  | |     |  |  |   |     |   |
Db  1051  VKKWKPPWEDDDEPSEEPVSAPEPEKKTPVLAKKAPTKPATKPDSEAAADPVSGPTSKDPK  1110

Qy   868  AT-------KPEMTTAKDKTTERDLRTTPETT----TAAPKMTKETATTTEKTTTESKITATT  919
          ||       |  |  |         |            | |  | | |  |   |
Db  1111  LSKKAPVEKKPKPTTDPDKDDLKPSPAKKPEKAPEPAAPKKWKPVWDDDPDPEPEADFTVPA  1170

Qy   920  TQVTSTTTQDTTPFKITTLTKTTILAPKVTVTTIKKTTITTTEIMNKPEETAKPKDRATNSKAT  979
                |  |        | | | |         ||
Db  1171  PSKKPDTEDPADPLG-------GPKTKDPK----------LNKKAPAEKPTEK----  1206

Qy   980  TPKPQKPKPTKAAPKPTSTKKPKTMPRVRKPK-------------TTPTPRKMTSTMP  1022
          || | | |  |  | | | ||      |
Db  1207  -PKPEVSKEPPKPTEPPKP-AAPKKWKPPWEDDDPDEPEADFTMPAPKKPDTEDP  1259
```

```
Qy    22  VSSQDLSSCAG---------------RCGEGYSRDATCNCDYNCQHYMECCCPDFKRVCTAAEVHN  71
          | :  |  | :               ||  : : | |  | | |       | ::  : |
Db   369  VNFDDRDNCLGALRPAQKTALQCAPVLARPVEVDCCS-----------------KHVC--AGYPT  413

Qy    72  PTSPPSSK-KAPPPSGCASQTIKSTTKRSPKPPNKKKTKVIESEEITEEHSVSENQESSS  130
          |    |:| ||    :       | |
Db   414  PGGPPSSSVPGKPPSVPGK------PAARAPMPTHITPPDVSPEPLPEPSPV-------  459

Qy   131  SSSSSSSSSTTIWKIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTPKKPPVVDEAGSGLDN  190
              :                                              |
Db   460  ------PAPAPMRMPTPLRSPPADEYI---------------------PTPPVPAKSPPG--  491

Qy   191  GDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKE  250
          |  :    | |   | |     |  |   |
Db   492  -----TSPPASRGA--PPLQAQPPAASSPPATPVKSSPPPAAV----------------  527

Qy   251  TTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPK  310
          |                                  |  | |
Db   528  -------------VLPPPAKTSPEPAPV-ASPPEAPVSSPPQVKSPP  562

Qy   311  EPTTPTTPKEPAPTTPKEPPTTTIKSAPTTPKEPAPTTTKSAPTTP-KEPAP------  362
          :| |  | |  | | |     |  |  | |  |   |  | |    | | |
Db   563  PPAPVASPPPMKSP--PPPARVASPPPLMKSPPPAPVASPPQLKSPPPPVLWLSTPS  620

Qy   363  TKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKE----PAP--TTP----KKPAPTTPKE  412
          | | | |  |  | | | |  | | | | |  |     |    ||       | | | |
Db   621  VKSPPPPVVASPPPPVKSPPPLAPVSSPSPPVKLPPLPAPGKSTPPPEEKPTPTPVK  680

Qy   413  PAPTTPKE-PTPTTPKEPAPTTTKEPAPTTPKEPAPTTPAPKPAPTTPKEPAPTTPKEPAPT  471
          | | | |   | | | | | | | |  |  | | | |    | | | | | | |
Db   681  SSPPPEKSLPPPTLTTSPPPQEKPTPPSSTSPSKPPPSPVETLPPPSKSSPPEEEVSSSPQ  740

Qy   472  TTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKKPAPTTPKEPAPTTPKETAPTTPKEPA  531
          |  | || | | | | | | | |  | | | |  | | | | | | | |  | | | |
Db   741  APKSSSP-----PAPVS--SPPPLKSSSPPEVPESSSPPPFPKSSPPLAPVSSPPVEKTSP  793

Qy   532  PTTPKKPAPTTPKEPAPTTTKKPAPTTPKEPAPTTPKKLTPTTP  591
          |  | | | | | | | | | | | | | | | | |  | |
Db   794  PPAPVSSPPPPTTPKSSSPPLAPVSSPPQVERTSPPPAPVSSPPPPTKSSPPLAPVSSPPPQVE  853
```

```
Query Match        11.2%;    Score 814;  DB 10;  Length 1480;
Best Local Similarity   28.8%;    Pred. No. 6.9e-45;
Matches   329; Conservative   28; Mismatches  541; Indels  212; Gaps  50;

Qy    592 EKLAPTTPEKPAPTTPEELAPTTPEEPTTPTPEEPAPTTPKAAAPNTPKEPAPTTPKEPA 651
Db    854 KTSPPPAPVSSSPPLEPKPSSPSSVSSSPPTTVKSSPPPAPLSSPPMTPKSSSPPAHVSSP 913

Qy    652 PTTPKEPAPTTPKETAPTTPKGTA-------------PTTLKEPAPTTPKPAPKELAPTTT 700
Db    914 PEAEKSSPPLAPISSPPSEPKSPSSSPPMVEKTSPPPATVSSPPPATVSSPPP---APVSS 970

Qy    701 KEPTSTTSDKPAPT-------TPKGTAPTTP--------KEPAPTTPKEPAPTTPK---GTAP 745
Db    971 PPPVVKSSPPPAPVSSPPPTTPKPLPPPAPVSSPPPPTPVSSPPPTTPKPLPPPP 1030

Qy    746 TTLKEPAPTTPKPAPKELAPTTIKGPTSTTSDKPAPTT--PKETAPTTPK----EPAPT 799
Db   1031 TPVSSPPPTYKPLPPP--APVSSPPPVVKSSPPPAPVSLPPPPTKPSPPRTRVSSPRPV 1087

Qy    800 TPKKPAPTTPETPPPTTSEVSTPTTIHKSPDESTPELSAEPTTPKALENSPKEPG 859
Db   1088 VKCCPPPTLVSSPPPAPKSLPPPIPVSSP----------PPEVKSSPPPTPVSSPPPAPK 1137

Qy    860 VPTTKTPAATKPEMTTTAKDKTTERDLRTTP-----ETTTAAPKMTKETATTTEKTTESK 914
Db   1138 SSPPPTPVVSSPPP------ELKSSPPPAPVSSPPPSSAPKSSPPPAPVSLPPPEVK 1185

Qy    915 ITATTTQVTSTTTQQDTTPFKITTLKTTLAPKVTTTKKITTTTELMNKPEETAKPKDRAT 974
Db   1186 SSPPPAPISSPPPAKSP-------------PPPA--------PMSSLPPVKSP 1219

Qy    975 NSKATTPKPQQKTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAM 1034
Db   1220 PPPAPVSSPPPPMKSPPPPAPISSPPPAP--VKPPSLPPPAPVSSPPPAV-------- 1267

Qy   1035 LQTTTRPNQTPNSKLVEVNPKSEDAGGAEGGETPHMLLRPHVFMPEVTPDMDYLPRVPNQG 1094
Db   1268 ---TSAP------------------------------PPAEALPPPSFND 1293

Qy   1095 IIINPMLSDE 1104
Db   1294 IILPIMANK 1303
```

```
RESULT   11
Q9LIE8

ID   Q9LIE8        PRELIMINARY;      PRT;  1480 AA.
AC   Q9LIE8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   STRAIN=COLUMBIA;
RC   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP001306; BAB03062.1; -.
DR   Interpro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE 1480 AA; 147153 MW; D1AC0C79F155E732 CRC64;
```

```
Qy     66 AAEVHNPTSPPSSSKRAPP----PSGGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSV 122
Db     51 AVKPHPHPKEPTIKPPPKRHPHPKPPTVK------PHPHPKPPTKPHPHPKPPPKPH-- 102

Qy    123 SENQESSSSSSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVK--DNKKNRIRKKKPPTKPPV 180
Db    103 -----------PHPKPPTIKPPPHPK----PRPHPKPPNVKPHPHPKPPTKPHPHPKPP- 146

Qy    181 VDEAGSGLDNGDFKVTTPDISTTQHNKVSTSPKITTAKPINPRPSLPPPNSDTISKETSLTV 240
Db    147 -------------------TKHH---PHPKPPTIKPPPKPPSVKPPPSTPK------- 175

Qy    241 NKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV--LAKPTPKAE 298
Db    176 ------PPTTNPPPSTPQPPPHIKPPPCTPTPPEVASSPPMATPPTQMPPIATP-PIAK 224

Qy    299 TTTTKGPALTTPKEPPT--TTP---KEPASTTPKEPTPPTTIKSAPTTPKEPAPTTT----K 349
Db    225 SPVATPPIATPPIATPPITIPPVATPPITTPPIANPPIIMPPIATPPVAAPPINPPISK 284

Qy    350 SAPTTPKEPAPTTT---------------KEPAPTTPKEPAPTTTKEPA------PTTTK 388
Db    285 PPVTTP----PTTTPPIAKPPIATPPISTPPAATPPIFITLPPAKPPVAISPIVT- 339

Qy    389 SAPTTP--KEPAPTTPKKPAP--TTPKEPAPTTPKEPTPTTPKEPAPTTPKEP-KEP 444
Db    340 -PPVTTPIAQPPVATPPVATPPIATPPSKSPISTPPISESPVATPPTATSPIKTP 398

Qy    445 APTAPKKPAPTTPKEPAPTTPKEPAPTTT---KEPSPTTPKEPAPTTTKSAPTTTKEPA 500
Db    399 PPAKPPVATPPIAKSPIATPPTPATPPVATPPIEKPPVATPPTTPPTTPPTAT--PPVAKPPV 455

Qy    501 PTTTKSAPTTPKEP--SPTTTKEPAPTTPKEPAP---TTPKKPAPTTPKEPAPTTPKEP- 554
Db    456 ETPPIATPPPIAKPPISTPPIISKPPVATPPAATPPIITPPTPVKPPVATPPLAIPPVAKPPV 515

Qy    555 -APTTTKKPAPTTAP--KEPAPTTPKEPAPTTT---------PKKLTP--TTPEKLAPTTPEKPA 603
Db    516 VTPPTATPPIATPPIAKSPVATPPATPPIAKPPVVTPPTTTPPTATPPVAKPPV 575

Qy    604 PTTPEELAPTTI--PEEPPTTPEEPAPTTPKAAAPNI----PKEPAPTTPKEPAPTTPKEPA 659
Db    576 ATPPIATPPTAKPPISTPPISKSPVATPPAATPPITTPPAKPPVATPPIATPPIAKPPV 635

Qy    660 PTTPKETAP----TTPKGTAPTT---LKEPAPTTPKKPAPKELAPTTTKEPISTTSDKPAPT 714
Db    636 ATPPIATPPIATSPVAKPPVATPPIKTPPPAKPPVAIPPIATPPVAKPPVAT--PPTAT 692

Qy    715 TPKGTAP--TTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPT 773
Db    693 PIATPPIATPPVTPPTATPPVATPPIAKPPTTI--PPTATPPVAMPPIATPPTAKPPI 750

Qy    774 ST---TSDKPAPTTPKETAPT--TPKEPAPTTPKEPAPTTP-----EEPPPTTSEVSTPTT 824
Db    751 ATPPIAIPPVAKPPVVTPPTTATPPIATPPIAKSPVATPPTATPPIAKPPVAIPPT 810

Qy    825 TKEPTTIH--KSPDESTPELSAEPTTPKALENSPKEPGVPTTKTPAATKPEMT--TTAKDKT 881
Db    811 TAPPTATPPVAKPPVATPPIATSPVAKPPVATPPIKTPPPAKPPVAIPPIATP 870

Qy    882 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQQDTTPFKITTLKTT 941
Db    871 ATPPIATPP--IAKPPVATPPIATPPIATSPVAKPPVAIPPIKTPPPAKPPVAIPPIATP 928

Qy    942 TLAPKVTTTKKT----IITTTELMNKPEETAKPKDRATNSKATTPKPQQKPTKAPKKPPSTK 997
Db    929 PVAKPPVATPPIATPPIATPPIATSPIATPPVVT---PPTATSPVATPPIAKPPTTTP--PTAT- 982
```

```
Qy    998 KPKTMPRVRKPKTTPTPRKMTSTMPELNP----------TSRIAE-------AMLQTTTRP 1041
Db    983 PPVAMPPIATP---PTAKPPVATPPIANPPVEKPPVVATPPTPIAKPPTVLPPPIAKPPVETSP 1039
Qy   1042 NQTPNSKL--VEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINP 1099
Db   1040 TATPPTATPPVAIPP------------VVKPPVAIPITKPPVVATPPVTNPPTAMPP 1084
Qy   1100 MLS 1102
Db   1085 IVT 1087
```

RESULT   12

ID   P70670
AC   P70670;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE   NACA.
DE   MUSCLE-SPECIFIC FORM GP220).
GN   NACA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96124450; PubMed=8698236;
RA   Yotov W.V., St-Arnaud R.;
RT   "Differential splicing-in of a proline-rich exon converts alphaNAC
RT   into a muscle-specific transcription factor.";
RL   Genes Dev. 10:1763-1772(1996).
DR   EMBL; U48364; AAB18734.1; -.
DR   EMBL; U48363; AAB18732.1; -.
MGD; MGI:106095; Naca.
DR   InterPro; IPR002715; NAC.
DR   InterPro; IPR003037; TS-N.
DR   Pfam; PF01849; NAC; 1.
DR   Pfam; PF02094; TS-N; 1.
SQ   SEQUENCE 2187 AA; 003646AA864DEBFD CRC64;

PRELIMINARY; PRT; 2187 AA.

Query Match   11.1%
Best Local Similarity 27.4%; Pred. No. 1.4e-44;
Matches 311; Conservative 128; Score 812; DB 11; Length 2187;
Mismatches 449; Indels 246; Gaps 51;

```
Qy     71 NPTSPPSSKKAPPPSGASQTIKSSTIKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSS 130
Db    746 SPT-PPSSKGAPVPSTGA--------------PPSPKGAPIVPTESSSISSKQVPAEILPSQ 792
Qy    131 SSSSSSSSSSTIWKIKSSKNSAANRELQKKLKYKDNKNRIKKKPTPKPPVVDEAGSGLDN 190
Db    793 KTPEVIASRLISAVQSPKVDPIMSD--------VPTSPKKTSATTAVPK----------- 833
Qy    191 GDFKVTTPDTSTTQHNK-----VSTSPKITTAKPINPRPSLPPPNSDTSKETSL--TVNKE 243
Db    834 DTSAFLSLKSVPAVTSLSPPKAPVAPSNEATIVPTEIPTSLKNALAAATPKE 885
Qy    244 T-------TVETKETTTINKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPPPKA 297
Db    886 TLATSIPKVTSPSPQKTPKSVSLKGAPAMTSKKAT---EIAASKDVSPSQ--FPKEVPLL 940
Qy    298 ETTTKGPALTTPKEPTTTPTTKEPASTTPKEPTTTIKSAPTTPKE-PAPTTTKSAPTTPK 356
Db    941 QHV----PPTSPPKSPVSDTLSGALTSPPPKGPPATLAETPTYPKKSSPKPAASKTPATPS 997
Qy    357 -------EPAPTTTKEPAPTTTKEPAPTTTKEPAPTT--TKSAPT---------- 392
Db    998 PEGVTAVPLEIPPCSKKAPKTAAPKESSATSSSKRAPKTAVSKEIPSKGVTAVPLEISLP 1057
Qy    393 -------TPKEPAPTTPKK-PAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKERAPTTPKE 443
```

```
Db   1058 LKETSKSATPGEKSASSPKRSPKTAGPKE---TPPGGVTAVPPEISLPPKETPQNATPNE 1114
Qy    444 RAPTARKKPAPTT--PKEPAP-----TTPKEPAPTTTKEPSPTTPK------EPAPTTKSA 492
Db   1115 SLAASSQKRSPKTSVPKETPPGGVTAMPLEIPSAPQKAPAKPKTAVPKQIPTPEDAVILAGS 1174
Qy    493 PTTTTKEPAPTTT-KSAPTTPK-------------EPSPTTTKEPAPTTTKEPAPTTTP---- 542
Db   1175 PLSPKRASKTAAPKEAPATPSVGTAVSGEISPSSPKKTSKTAAPKENSATLPPKRSPKTA 1234
Qy    543 -PKE-------------------------------PAPTTTKEPAP------------- 582
Db   1235 APKETPATSSEGVTAVPSEISPSSPPTPASKGVPVTLTPGGAPNALAE-SPASPKKVPKTA 1293
Qy    583 PKKLTPTTPEKLAPTTPK--PAPTTPEELAPTTPEEPTPTT------PEEPAPTTPKA 633
Db   1294 APEETSTTP-----SPQKIPKVAGPKEASATPPSKKTPKTAVPKETSAPSEGVTAVPLE 1347
Qy    634 AAPNTPKEPAPTTPKE-PAPTTPKEPAPTTPKETAPTTTKGTAPTTLKEPAPTTP----- 687
Db   1348 IPPSPKRAPKTAAPKETPAPS--PEGATTAPVQIPPSPSRKGSKKAGSKE-TPTTSPEGV 1404
Qy    688 ------------KKPAPKELAPTTTKEPTS---------TTSDKPAPTTPKGTAPTTPKE 726
Db   1405 TAAPLEIPISSKKTSKMASPKETLVTPSSKKLSQTVGPKETSLEGATAVPLEIPPSHKKA 1464
Qy    727 PAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTTKGPTSTTSDKPAPTTP 785
Db   1465 PKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-APKTAAPPSER-VTTVPPEKPA-TPQ 1519
Qy    786 KETAPTTPKEPAPTTPKEPAPTTPETP------PPTTSEVSTPTTTK-----EPTTHKS 834
Db   1520 KASGTTASKVPVPAETQEVAVSSRETPVTPAVPPVVKNPSSHKKTSKITELKEAPATLPPS 1579
Qy    835 PDESTPELSAEPTPKALENSPKE-PGVPTTKTPAATKPEMTTTAAKDKTTERDLRTTPETT 893
Db   1580 PTKSPKIPSSKKAPRT--SAPKEFPASPSIK-PVTT-SLAQTAPPSLQKAPSTTIPKEN 1634
Qy    894 TAAPKMTKETATTTESKITATTTQVTSTTTQODTTPFKITTLKTTTLAPKVTTTKKT 953
Db   1635 LAAPAV-----LPVSSKSPAAPARASASLSPATAAPQQTAPKEANTTIPSCKKAAATETPIET 1690
Qy    954 ITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRKPKTTPT 1013
Db   1691 STAPSLEGAPKETSE----TSVSKVLMSSP-----PKKASSSKRASTLP---------- 1730
Qy   1014 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETP 1067
Db   1731 ----ATLIPSLKEASVLS--------PTAISSGKDSHISPVS-DACSTGTTTP 1770
```

RESULT   13

ID   Q41805
AC   Q41805;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z34465; CAA84230.1; -.
DR   Mendel; 14346; Zeana.2368; 14346.
DR   InterPro; IPR001611; LRR.

PRELIMINARY; PRT; 1188 AA.

```
DR      InterPro; IPR003592; LRR_out.
Pfam;   PF00560; LRR; 3.
DR      SMART; SM00370; LRR; 4.
SMART;
KW      Signal.
FT      SIGNAL
SQ      SEQUENCE   1188 AA;  120980 MW;  2C77C7F8D7130149 CRC64;
                   1               27        POTENTIAL.
```

```
RESULT  14
Db      Q9VEL9
ID      Q9VEL9;     PRELIMINARY;    PRT;   2112 AA.
AC      Q9VEL9;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE      CG4090 PROTEIN.
```

```
GN      CG4090.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BERKELEY;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA      Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport A.R., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA      Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).
DR      EMBL; AE003716; AAF55402.1; -.
DR      FlyBase; FBgn0038492; CG4090.
DR      InterPro; IPR002557; Chitin_binding.
DR      InterPro; IPR005061; EGF-like.
DR      Pfam; PF01607; Chitin_bind_2; 10.
RT      SMART; SM00494; ChtBD2; 11.
DR      PROSITE; PS00022; EGF_1; 1.
DR      SEQUENCE   2112 AA;  B91018E5551A5D36 CRC64;
SQ
```

```
Query Match
Best Local Similarity    10.7%;    Score 784;   DB 10;    Length 1188;
Matches 230;   Conservative 59;   Pred. No. 4.9e-43;
                29.9%;                    Mismatches 359;  Indels 122;  Gaps 33;
```

```
Qy    174  PTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL----PPN  229
           |||                   :|.||.|     :|::||||::|||||.|||      | |
Db    1037 KVVSQGGSSSNSTSNSSSSSNNSGSSSNSGSSSSS-----SSSNS------------    1076

Qy    230  SDTSKETSLTVNKETTVETKETTTNKQTTDTGKEKTTSAKETQSIEKTTSAKDLAPTSKV  289
Db    987  NCSKFYRCVDNGKGGFTKVSF---TCPPNTLWDPEANSCNIPDQIQTK-------PLKCK 1036

Qy    290  -----LAKPTPKAETTKGPALTTPKEPPTTPKEPPTTPKEPASTTPKEPTPTTPKEPA  344
Db    943  ...

Qy    345  PTTTKSAPTTP---KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTP-KERAPTT  401

Qy    402  PKKPAPTTPKEPAPTT-----PKE--PTPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKP  452

Qy    453  APTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK  512

Qy    513  EPSPTTPKEPAPTTPKPAAAPNTPKE-PAPTTPKEPAPTTPKKPAPTTPKRPAPTAPKEPAP  572

Qy    573  TTPKETAP----TTP---KKLTPTTPEKLAPTTPEKPAP----TTPEELAPTTPEEPT  619

Qy    620  PPAPLSSPPPAPQVKSSPPPVQVSSP-PPAPKSSPPLAPVSSPPQVEKTSPPPAPLSSPP  659

Qy    660  PAKSTPPPEKYPTPPTTSVKSSPPPEKSLPPTLIPSPPPQEKPTPPSTPSKP-PSSPEKP  718

Qy    679  LKE--PA----PTTPKK--PAPKELAPTTTKE--PTSTTSDKPAPTTPKGTAPTTP  724

Qy    725  -KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTT  783

Qy    769  PPAPLSSPPSAPAPQVKSSPPPVQVSSP-PPAPKSSPPLAPVSSPPQVEKTSPPPAPLSSPP  827

Qy    784  -TPKETAPTTTPKEPAPTTPKPAPTTPETPTPTTPSEVSTPTTTKEPTTIHKSPDESTPEL  842

Qy    828  LAPKSSPPPHVVVSSPPPVVKSSPPPAPVSSPPLTEKPASPPAHVSSPPPEVVKPSTP--PA  885

Qy    843  SAEPTPKALENSPKEPGVPTTKPAATKPEMTTTAKIDKTTERDLRTTPET  892

Qy    886  PTTVISP-PSEPKSSPPPTPVSLPPPIVKSSPPPAMVSSP-PMTKSSPPPVVVSSPPPT  943

Qy    944  VKSSPPPPAPVSSPPAIPKSSPPPAPVNLPPPEEVKSSPPPTPVSSPPPA--PKSSPPPAP 1000

Qy    1001 MSSPPPPPEVKSPPPPAPVKSSPPPPVKSSPPPPAPVSSPP----PPVKSPPPPAPVSSPPPPV 1057

Qy    1058 KSPPPPAPISSSPPPPVVKSSPPPPPPAPVS--SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS 1115

Qy    1116 SPPPAP-------VKPSLP-PPAPVSSPPPVVTPAPPKKEEQSLPPPAES 1158
```

```
Query Match
Best Local Similarity    10.7%;    Score 781;   DB 5;    Length 2112;
Matches 309;   Conservative 98;   Pred. No. 1.4e-42;
                27.3%;                    Mismatches 367;  Indels 356;  Gaps 48;
```

```
Qy    49   NCQHYMCCCPDFKRVCTAAAEVHNPTSPPSSKKAPPSGASQTIKSTTKRSPKPPNKKKTK  108
           ||  |.:     |  |  | :|||      |:.|    |:||:||  |    ||
Db    987  NCSKFYRCVDNGKGGFTKVSF---TCPPNTLWDPEANSCNIPDQIQTK-------PLKCK 1036

Qy    109  KVT----ESEEITEEHSVSENEQESSSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKLKVKD  164
           ||:      |  |         |.||||||||||||||||            .||||
Db    1037 KVVSQGGSSSNSTSNSSSSSNNSGSSSNSGSSSSS-----SSSNS------------ 1076

Qy    165  NKKNRTKKKPTPKPPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRP  224
           |: |      .|||||||||||||||||||||||||||||||||||||||||||:|:||
Db    1077 ------------GSSSNTGSSSNGASSSGGSSSNQGSSS-------------------- 1103
```

```
RESULT  15
Q9XDH2
ID   Q9XDH2          PRELIMINARY;      PRT;    763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH MUCIN HOMOLOG.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang Y., Moreno C., Singh M.;
RT   "Cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003882; Pistil_extensin.
DR   PRINTS; PR01222; ATROPHIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE  763 AA;  75034 MW;  39168EC45A5916F8 CRC64;
```

```
Query Match       9.8%;     Score 715;     DB 2;     Length 763;
Best Local Similarity   28.2%;
Matches  248;  Conservative  39;  Mismatches  361;  Indels 232;  Gaps  39;
```

```
Qy  225 SLPPNSDTSKETSLIVNKETTVVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAK-DL  283
Db 1104 -----NSGSSGSNSSGNQSTS----------SSTSSSSSSSNNNQGSSSSSSSSSKPNP  1154

Qy  284 APTSKVLAKPTPKAETTTK----------------------------------GP----  304
Db 1155 SETCKVNGQFIGDRSDCAKFYRCVDNDRGGEFNNVPFSCGPGTVWDAQMQACNHAWAVKEC 1214

Qy  305 -ALTTPKEPTTPTPKEPASTTTPKEPTTTIKSAPTTTPKEPAPTTTK              363
Db 1215 GGTAPPTTSPTTTSR-PTTASRSRPSDQTSTSRPTGP----PTTARPVTARPTTSSPPTTA 1269

Qy  364 KEPAPTTPKEPAPTT-------TKEP----APT----                        385
Db 1270 SSSQTTSPVTQAPNTDGKCRSEGFMADPNNCSKFYRCVRNNKGGFTSIPFQCGAGTVWDQ  1329

Qy  386 ---TTKSAPTTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTTT             433
Db 1330 DLQTCNHNFNNCSTGTESTTPKPPC--EPATNGTTATSTSSTTTP--PPTTDLPPTSTT  1385

Qy  434 KEPAPTTPKEPAPTTAPKKPAPTTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAP        493
Db 1386 GLP-PTTTTELPPT--------TTTDLPPTTTTRLPPTTTTSLPPTTTTGLPPTTTTGAQP 1437

Qy  494 TTTKEPAPTTTIKSAPTTTPK---EPSPTTTKEPAP                          524
Db 1438 TTTTLSSETETTSTVTTSPESTTQPPSTTTMKPLPAGTECTGEGYMADPEDCRKYRCINA  1497

Qy  525 -------TTPK------------------EPAP----TTPKKPAPTTP---KEPAPTTPKE  553
Db 1498 GASYRKYNFTCPKGIGWNEEVQTCDYVENIPRCSKLPAEPTTTTTPSEESKDPGSTTPQST 1557

Qy  554 PAPTTTKKPAPTTAPKEPAPTTPKETAPTTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPT 613
Db 1558 DEPTTVTKPI-TKPTE-EPSTEKPQKPTTQYPEKPTTTEE--PERKPQKPT-TTEYPQKPT  1612

Qy  614 TPEEPPTTPTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTT-PKEPAPTTPKETA----  667
Db 1613 TTEEPPPEKPQKPT--TTEYPQKPTTTEEPPTPEKPQKPTTTEEPTTTSIPGYN        1671

Qy  668 PTTPK--GTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPK  725
Db 1672 PTTTSVPGYNPTTTPIPVETTTSTPGYK----PITTGEPITTTT--LPSTTTDAIQEPPTSK 1727

Qy  726 EPAPTTPKE-PAPTTPKGTAPTTLIKEPAPTTPKK------PAPKELA-----PTTTKGPT  773
Db 1728 KPEPTTTESPESSTPEGSVTTLQPEPQPNYNCSSEGFPDPAEDCSRYYRCVDAAAKNGKY  1787

Qy  774 STTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTTPETEPPPTTISEVSTPPTTTKEPTTIHK 833
Db 1788 QVYAFKCGKGTVWDTSTETCNYADQVSGN--------CSSQTTTPGTTTTEPGTTES      1836

Qy  834 SPDESTPELSAEPPPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETT  893
Db 1837 TTSSGKP----ETTSKAPEN-------TTTWA---PETTTT----SSPETTT            1869

Qy  894 TAAPKMTKETATTTTEKTTESKITATTTQVTSTTTQQDTTPFKITTTLKTTTIAPKVTTTKKT 953
Db 1870 TT----VASETTTTTSGT-----------TTTATTPETTTKPPKP--ETTTIAGETSTSKS  1913

Qy  954 ITTTEIMNKPEETAKPKDRATNSKATTTPKPQKPTKAPKKPTSTKKPKTMP          1003
Db 1914 PTTTE---------------------------SPAPSTNTSAP--------CPETGP      1935
```

```
Qy  293 PTTPKAETTTKGPALTTTPKEPTTTPKEPASTTTPKEPTTTPKEPAPTTTKSAP       352
Db    3 PVP------APRALAPLPPAPPAPAEPKSKPPFPPAP----------PAPPCWMLVSAAP  46

Qy  353 TTPKEPAPTTPKKPAPTTTPKEPAPTTTPKEPTTTKSAPTTTPKEPAPTTTPKKPAPTTTPKE 412
Db   47 PCP--PAP----PAPPKPKSKAPPPVPPAPPAPARELAPLP--PAP------PEAPRE      90

Qy  413 PAPTTPKEPTPTT----PKEPAPTTKEPAPTTPK-EPAPTTAPK-KPAPTTPKEPAPTTPK  466
Db   91 SRPALPPCPPCPPPVVIPDPPEPAAPVPPAPNSPPPPPPFFPPAPKEVPAPPVP--PVPNSPP 148

Qy  467 EP--APTTTKEPSPTTPKEPAPTTTKSAPTTTPKEPAPTTTKEPSPTTPKEPAP         524
Db  149 FPPFPPAALNPPAP-----PAPPLANSPPLPPAPPPAGT--PPAAPWPPVPAAPKSKPA    201

Qy  525 TTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAP-----TTPKKPAPTTAPKEPAPTTPKETAP 580
Db  202 SPPRPPAP----PMPATPMEFPPLPVPVPPDPPISKETPAPPAPPIPPAPAPVPIPPVPPLP  256

Qy  581 TTPKKLTPTTPEKLAP--------TTPEKPAPTTPEELAPTTTEEPTTTTPEEPTTPTT-   630
Db  257 PVPNKIPPAPP---APPVAVAVALVAPCPPLPPLPNNHPPAPPAAPVPGVPLAPLPNSHP   313

Qy  631 ---PKAAAPNTPKEPAP                                             644
Db  314 PAPPSSAPVPGVPLAPLPISGRPVSVWKGSPTTLLSTFCCRVCSGEVLAGALNPSRSPL   373

Qy  645 --TTPKEPAPTTPKEPAPTTPKETA----PTTPKGTA---PTTLKEPAPTTP-KKPAPKE  694
Db  374 TTTTPALPAPIPLPPLPLPINTAVPPTTPLPPVTALAPPLPPLAPLPISPGVPPAP--     431

Qy  695 LAPTTTKEPTTTISDKPAPTTPKGTAPTTPKEPA-----PTTPKEPAPTTPKGTAPTTTL  748
Db  432 --PIPPGKPWTTPLAPAPPEPRK-TVPVLPPGCPPSEKPNPPAPPEPPEPKSSPALPP    488

Qy  749 KEPAPTTP----KKPAPKELAPTTTKGP-TSTTSDKPAPTTPKETT-APTTTKEPAPTTP  801
Db  489 APPAPSMPSAVRVPPSSPPIPPAPPAARASMPALPPAPSPPAATRLCPPLPPSPPAPNSP   548

Qy  802 KKPAPTTPETPPTPPPTTSEVSTPTTTTKEPTTIHKSP------DESTPELSAEPTPKA--LENSP 855
Db  549 --PAPPAPPTPKLLS--ANPCPPVPPAPNRPPAPPAPPAPPELRAPPDPDTTPPVANSP   604

Qy  856 KEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTIESKI 915
```

```
Db   605 --PAPPAPPAPPSALPFVNPPA---------------------PPTPAAPK-------------SRPAL 637

Qy   916 TATTTQVISTTTQDTTPFKITTTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATN 975
             :   |||
Db   638 PAAPPAPPAPPVRATTP-------                   ----PPAPPAPPAPN 665

Qy   976 SKATTPKPQKP-----TKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTM-----PELNP 1026
             |   |  | : |   |              | | |               |:    | | |
Db   666 SMALPPAPPDPPIPLLATPPAPPAPPLPMSPPAPPLPPAAPDPPAPPLTINQPPSPPLAP 725

Qy  1027 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET 1066
             | |    ||   : |:           : |:|
Db   726 VPGAPLAPLPINGRPVFARKNSLI-------GSSSGDT 756
```

Search completed: April 26, 2002, 16:29:54
Job time: 609 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:21:42 ; Search time 153.5 Seconds
                     (without alignments)
                     1315.022 Million cell updates/sec

Title: US-09-556-246-1_COPY_25_1404
Perfect score: 7410
Sequence: 1 QDLSSCAGRCGEGYSRDATC..........ARAITTRSCQTLSKVWYNCP 1380
Scoring table: BLOSUM62
                     Gapop 10.0 , Gapext 0.5
Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Post-processing: Minimum Match 0%
                     Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
                     Listing first 45 summaries

Database :   SPTREMBL_17:

                     1: sp_archea:*
                     2: sp_bacteria:*
                     3: sp_fungi:*
                     4: sp_human:*
                     5: sp_invertebrate:*
                     6: sp_mammal:*
                     7: sp_mhc:*
                     8: sp_organelle:*
                     9: sp_phage:*
                    10: sp_plant:*
                    11: sp_rodent:*
                    12: sp_virus:*
                    13: sp_vertebrate:*
                    14: sp_unclassified:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	%Query Match	Length	DB	ID	Description
1	7410	100.0	1404	4	Q92954	Q92954 homo sapien
2	7381	99.6	1404	4	Q92954	Q92954 homo sapien
3	3913	52.8	1054	11	Q9BX49	Q9bx49 homo sapien
4	1713.5	23.1	401	6	Q9JM99	Q9jm99 mus musculu
5	982.5	13.3	1079	5	O77765	O77765 bos taurus
6	946.5	12.8	1049	5	Q9N4S7	Q9n4s7 caenorhabdi
7	902	12.2	1795	5	Q9I7S1	Q9i7s1 drosophila
8	863.5	11.7	1489	10	O76894	O76894 drosophila
9	846	11.4	1274	5	O96449	O96449 phytophthor
10	816	11.0	1315	10	Q20007	Q20007 caenorhabdi
11	812	11.0	2187	11	Q9SPM0	Q9spm0 zea mays (m
12	811.5	11.0	1480	10	P70670	P70670 mus musculu
13	784	10.6	1188	10	Q9LIE8	Q9lie8 arabidopsis
14	782.5	10.6	2112	10	Q41805	Q41805 zea mays (m
15	715	9.6	763	2	Q9VEL9	Q9vel9 drosophila
16	709.5	9.6	677	5	Q9XDH2	Q9xdh2 mycobacteri
17	707.5	9.5	555	10	Q9N435	Q9n435 caenorhabdi
18	703.5	9.5	4880	11	Q9FPQ6	Q9fpq6 chlamydomon
19	703.5	9.5	5085	11	Q9JLT1	Q9jlt1 rattus norv
19					Q9JKS6	Q9jks6 rattus norv

ALIGNMENTS

RESULT   1
ID   Q92954   PRELIMINARY;   PRT;   1404 AA.
AC   Q92954;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEGAKARYOCYTE STIMULATING FACTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA   Jacobs K., Turner K.;
RL   (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL   Mosher D.F. (eds.);
RL   Biology of vitronectins and their receptors., pp.45-52,
RL   Elsevier Science Publishers B.V. (1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Blood 78:279-279(1991).
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RL   EMBL; U70136; AAB09089.1; [3]
DR   EMBL; U70136; AAB09089.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   InterPro; IPR002400; GF_cysknot.
DR   Pfam; PF00045; Hemopexin; 2.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR00438; GFCYSKNOT.

20	695	9.4	1151	13	O57580	O57580 gallus gall
21	676.5	9.1	1229	5	Q94185	Q94185 caenorhabdi
22	674.5	9.1	2089	4	O14676	O14676 homo sapien
23	674.5	9.1	2768	5	Q9VC00	Q9vc00 drosophila
24	669	9.0	2284	5	Q9VPG1	Q9vpg1 drosophila
25	666.5	9.0	7962	4	Q10465	Q10465 homo sapien
26	665	9.0	4833	11	Q9QYX6	Q9qyx6 mus musculu
27	665	9.0	5038	11	Q9QYX7	Q9qyx7 mus musculu
28	663	8.9	3570	4	Q99552	Q99552 homo sapien
29	658.5	8.9	3507	10	Q23587	Q23587 caenorhabdi
30	653.5	8.8	1514	5	Q9GUW7	Q9guw7 leishmania
31	643.5	8.7	990	13	Q91803	Q91803 xenopus lae
32	636	8.5	6632	5	O17362	O17362 caenorhabdi
33	633	8.5	489	10	O41707	O41707 vigna ungui
34	632	8.5	761	10	Q92QI0	Q92qi0 arabidopsis
35	631	8.5	2344	5	Q9N3Y8	Q9n3y8 caenorhabdi
36	629	8.5	6642	10	O01761	O01761 caenorhabdi
37	628.5	8.5	1612	5	Q9VQ2	Q9vq2 drosophila
38	627.5	8.5	971	5	Q9XVS4	Q9xvs4 caenorhabdi
39	622.5	8.4	839	2	Q9RX57	Q9rx57 deinococcus
40	607.5	8.2	801	5	Q23635	Q23635 caenorhabdi
41	607.5	8.2	924	12	Q99307	Q99307 epstein-bar
42	605	8.2	379	5	Q27929	Q27929 drosophila
43	600.5	8.1	1893	5	Q9NKC9	Q9nkc9 drosophila
44	598.5	8.1	409	10	Q9SBM1	Q9sbm1 volvox cart
45	591.5	8.0	2232	5	P91365	P91365 caenorhabdi

```
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
SQ   SEQUENCE 1404 AA;  151090 MW;  AABD7AD19B35F4F6 CRC64;

Query Match            100.0%;  Score 7410;  DB 4;  Length 1404;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 1380; Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDFKRVCTAELSCKGRCFESFERGREC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDFKRVCTAELSCKGRCFESFERGREC 84

Qy     61 DCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAPPSGASQTIKSTTKRSPKPNKKK 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 DCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAPPSGASQTIKSTTKRSPKPNKKK 144

Qy    121 TKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 TKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK 204

Qy    181 KNRTKKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 KNRTKKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL 264

Qy    241 PPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPT 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 PPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPT 324

Qy    301 SKVLAKPTPKAETTTKGPALTTPKEPPTTTPKEPPTTTPKEPPTTTIKSAPTTPKEPAPT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 SKVLAKPTPKAETTTKGPALTTPKEPPTTTPKEPPTTTPKEPPTTTIKSAPTTPKEPAPT 384

Qy    361 TTKSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTKKPA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385 TTKSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTKKPA 444

Qy    421 PTTPKEPAPTTTPKEPTTTPKEPTTTPKEPAPTTTKEPAPTTTKKPAPTTPKEPAPTTPK 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 PTTPKEPAPTTTPKEPTTTPKEPTTTPKEPAPTTTKEPAPTTTKKPAPTTPKEPAPTTPK 504

Qy    481 EPAPTTTTKEPSPTTPKEPAPTTTKSAPTTTKSAPTTTKEPAPTTTKSAPTTPKEPAPTT 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 EPAPTTTTKEPSPTTPKEPAPTTTKSAPTTTKSAPTTTKEPAPTTTKSAPTTPKEPAPTT 564

Qy    541 PKEPAPTTPKKPAPTTTKKPAPTTPKEPAPTTPKETAPTTTKKL 600
          ||||||||||||||||||||||||||||||||||||||||||||
Db    565 PKEPAPTTPKKPAPTTTKKPAPTTPKEPAPTTPKETAPTTTKKL 624

Qy    601 TPTTPKEKLAPTTPKKPAPTTPEELAPTTPEEPTTTPEEPAPTTTPKAAAPNTPKEPAPTT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625 TPTTPEKLAPTTPEKLAPTTPEELAPTTTPEEPTTTPEEPAPTTTKAAAPNTPKEPAPTT 684

Qy    661 PKEPAPTTPKEPAPTTPKKPAPTTPKRTAPTTPKGTAPTTLKEPAPTTTKEPTST 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    685 PKEPAPTTPKEPAPTTPKKPAPKELAPTTLKEPAPTTTKEPTST 744

Qy    721 TSDKPAPTTPKGTAPTTPKEPAPTTPKRPAPKELAP 780
          ||||||||||||||||||||||||||||||||||||
Db    745 TSDKPAPTTPKGTAPTTPKEPAPTTPKRPAPKELAP 804

Qy    781 TTTKGPTSTTSDKRPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPSEVSTPTTTK 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805 TTTKGPTSTTSDKPAPTTPKETAPTTPKETAPTTTPKPTPPTTTSEVSTPTTTK 864

Qy    841 EPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDL 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    865 EPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDL 924

Qy    901 RTTPETTTAAPKMTKETATTTEKTTTESKITATTTQVTSTTTQDTTPFKITTTLKTTTLAPK 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    925 RTTPETTTAAPKMTKETATTTEKTTTESKITATTTQVTSTTTQDTTPFKITTTLKTTTLAPK 984
```

```
RESULT 2
Q9BX49

ID   Q9BX49          PRELIMINARY;      PRT;   1404 AA.
AC   Q9BX49;
DT   01-JUN-2001 (TREMBLrel. 17, Created)
DT   01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).
GN   BG174L6.2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133553; CAC36090.1; -.
SQ   SEQUENCE 1404 AA;  782A1174B3FDEE5 CRC64;

Query Match            99.6%;   Score 7381;  DB 4;  Length 1404;
Best Local Similarity  99.7%;   Pred. No. 0;
Matches 1376; Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

Qy      1 QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDFKRVCTAELSCKGRCFESFERGREC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDFKRVCTAELSCKGRCFESFERGREC 84

Qy     61 DCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAPPSGASQTIKSTTKRSPKPNKKK 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 DCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAPPSGASQTIKSTTKRSPKPNKKK 144

Qy    121 TKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 TKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK 204

Qy    181 KNRTKKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 KNRTKKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL 264

Qy   1321 GLPNVVTSAISLPNIRKPDGYDYYAFSKDQYINIDVPSRTARAITTRSGQTLSKVVWNCP 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1345 GLPNVVTSAISLPNIRKPDGYDYYAFSKDQYINIDVPSRTARAITTRSGQTLSKVVWNCP 1404
```

```
Qy   241  PPNSDTSKETSLIVNKETTVETKETTTNKQTSTDGKERTTSAKETQSIEKTSAKDLAPT  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265  PPNSDTSKETSLIVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPT  324

Qy   301  SKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTTIKSAPTTPKEPAPT  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   325  SKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPT  384

Qy   361  TTKSAPTTPKEPAPTTPKEPAPTTPKEPTPTTPKEPAPTTTPKEPAPTTPKKEPAPTT  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   385  TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTPKEPAPTTPKKEPAPTT  444

Qy   421  PTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKKPAPTTPKKPAPTTPKKEPAPTTPKK  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   445  PTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKKPAPTTPKKEPAPTTPKK  504

Qy   481  EPAPTTTTKEPSPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   505  EPAPTTKEPSPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTT  564

Qy   541  PKEPAPTTTKKPAPTTPKEPAPTTPKEPAPTAPKEPAPTTPKETAPTTPKKL  600
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   565  PKEPAPTTPKKPAPTTPKEPAPTTPKRPAPTTPKEPAPTTPKETAPTTPKKL  624

Qy   601  TPTTPEKLAPTTPKEPAPTTPEELAPTTPEEPTPTTPKAAAAPNTPKEPAPTT  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   625  TPTTPEKLAPTTPKEPAPTTPEELAPTTPEEPTPTTPKAAAAPNTPKEPAPTT  684

Qy   661  PKEPAPTTPKEPAPTTPKETAPTTLKEPAPTTTTKEPTST  720
          |||||||||||||||||||||||||||||||||||
Db   685  PKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTTTKEPTST  744

Qy   721  TSDKPAPTTPKGTAPTTPKGTAPTTPKGTAPTTLKEPAPTTPKRPAPKELAP  780
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   745  TCDKPAPTTPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKRPAPKELAP  804

Qy   781  TTTTKGPTSTTSDKPAPTTPKETAPTTPETPPPPTTPEVSTPTTTK  840
          ||||||||||||||||||||||||||||||||||||||||
Db   805  TTTKGPTSTTSDKPAPTTPKETAPTTPKKPAPTTPEPTPTTPTSEVSTPTTK  864

Qy   841  EPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTPKTPAATKPEMTTTAKDKTTERDL  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   865  EPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKITPAATKPEMTTTAKDKTTERDL  924

Qy   901  RTTPETTTIAAPRMTKEFATTTESKITATTTQVTSTTQDTTPFKITTLTKTTTLLAPK  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   925  RTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTQQTTPKITTLKTTTLLAPK  984

Qy   961  VTTTKKIITTTEIMNKEEETARPKDRATNSKAITTPKQKPTKAPKKPTSTKKPKTMPRVR  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   985  VTTTKKIITTTEIMNKEEETAKPKDRAITNSKAITTPKQKPTKAPKKEPTSTKKPKTMPRVR  1044

Qy  1021  KPKTTPTPPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET  1080
          ||||||||||||||||||||||||||||||||||||||||||||
Db  1045  KPKTTPTPPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET  1104

Qy  1081  PHMLLRPHVFMPEVTPDMDYLPRVPVPNQGIINPMLSDETNICNGKPVDGLTTLRNGTLVA  1140
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  1105  PHMLLRPHVFMPEVTPDMDYLPRVPVPNQGIIINPMLSDETNICNGKPVDGLTTLRNGTLVA  1164

Qy  1141  FRGHYFWMLSPSPPSPSPARRITEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIK  1200
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1165  FRGHYFWMLSPSPPSPSPARRITEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIK  1224

Qy  1201  DAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYFFKRGGSIQQVIYKQEPVQKCPGR  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1225  DAGYPKRIFKGFGGLTGQIVAALSTAKYKNWPESVYFFKRGGSIQQVIYKQEPVQKCPGR  1284

Qy  1261  RPALNPVYGEMTQVRRRFERAIGPSQOTHIRIQYSPARLAYQDKGVLHNEVKVSILWR  1320
          |||||||||||||| |||||||| |||||
Db  1285  RPALNPVYGETTQVRRRFERAIGPSQOTHIRIQYSPARLAYQDKGVLHNEVKVSILWR  1344

Qy  1321  GLPNVVTSAISLPNIRKPDGYDYYAFSKDQYNIDVPSRIARAITTRSGQTLSKVWYNCP  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1345  GLPNVVTSAISLPNIRKPDGYDYYAFSKDQYNIDVPSRIARAITTRSGQTLSKVWYNCP  1404
```

```
Query Match          52.8%;   Score 3913;   DB 11;   Length 1054;
Best Local Similarity   57.1%;   Pred. No. 4.6e-245;
Matches   795;   Conservative   57;   Mismatches 166;   Indels 374;   Gaps 25;

Qy     1  QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDFKRVCTAELSCKGRCFESFERGREC  60
          |:|||||||||||||||||||| |||| ||:||:||||||||| || |||:|||||:||
Db    25  QDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDFKRVCSPELSCKGRCFESFARGREC  84

Qy    61  DCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKRAPPSGASQTIKSTTKRSRPKPPNKKK  120
          |||| |||| ||||||:||||:||||:||||:|||||||:||||||||||:||:||||||
Db    85  DCDSQCKCQVGKCCADYDSFCEEVHNSTS-PSSKTAPTPAGASDIIKSTTKRSRSPKSPT-TR  142

Qy   121  TKVIESEITEEHSVSENQESSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK  180
          ||:||||:||:|||||||||||||||||||||||||:||:||:|||||||:||||
Db   143  TIKVVESEELTEEHSDSENQE--SSSSSSSSTIRKIKSSKNS-ANRELQKNPVKDNK  199

Qy   181  KNRTKKKPTPKPPVVDEAGSGLDNGDFKVT--TPDTSTTQHNKVSTSPKITTAKPINPRP  238
          ||| |||||||||||||||||||||||| |: :|||||||||||:|||||||:||||:
Db   200  KNTPKKKNPEPPAVDEAGSGLDNGEFKLTPPPDPDPPTTPHSKVATSPKTTAAKPVTPKP  259

Qy   239  SLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLA  298
          ||||||||||||||||||||||||||||||||||||||||||||||||||:||:|||||
Db   260  SLAPNSEISKEASLASNKETTVETKETTATNKQSSA--SKKKTTSVKETRSAEKTSDKDV-  317

Qy   299  PTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPA  358
          ||| ||||||||||||||| |||| |:||||||||||||||| ||||||||||||||||
Db   318  PTS-----------------EPTSTTPK------NSAPTTKKPV  336

Qy   359  PTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTIKSAPTTPKEPAPTTPKK  418
          ||||||||||||||||||||||||||||| |||||| |::|||||||||:|||||||||:
Db   337  -TTTKESKPLP----------------------LPQEPEPTTAKEPPTTKKPEPPTTKRE  381

Qy   419  PAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT  478
```

```
Db  382 PEPTTPKEPAPTTPKKEPSPTTPKEPAPTTKSAPTTTKEPAEPTTPKEPEPTTPKEPEPTT  441
        | |||||  |||||| ||||||||  |||| |||||| |||||||||||||||  |||| |||
Qy  479 PKEPAPTTPKEPSPTTPKEPAPTTTKEPEPTTPKEPEPTTPKEPEPTTPKEPEPTT       538

Db  442 TKEPESTTRKEEPTTPKEP------------------EPTTPKEPEPTTLKEPEP        479
        | |||  |||||||||||                  ||||||||||||| ||||
Qy  539 TTPKEPAPTTPKKPAPTTPKEPAPTTPKKPAPTAPKEPAPTTPKETAPTTPK          598

Db  480 TTPKEPEPTTPKEPEPTTPKKPAPTTTKEPEPTTPKEP-----------TT           512
        |||||||  |||| ||||||| || |||| |  | ||           ||
Qy  599 KLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTPTPEEPAPTTPKAAAANTPKEPAP  658

Db  513 TTPKEPEPTTPKEPEPTTP------------------------------TPKEPEP      519
        | |||||||||||||||||                            ||||||||
Qy  659 TTPKEPAPTTPKEPAPTTPKEIAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTPKPT  718

Db  520 TTPKEPEPTTPKGTAPTTP------------------------KEPEPTTPKKP        549
        |||||||||||| ||||||                      |||||||||||
Qy  719 STTSDKPAPTTPKGTAPTTPKEPAPTTLKEPAPTTPKKPAPKEL                 778

Db  550 -----------------------EPTTPKEPVPTTPKEPETTP--------KEPEPTTPKEPEP  583
                              ||||||||||||||||||       |||||||||| |
Qy  779 APTTKGPTSTTSDKPAPTTPKEIAPTTPKKPAPTTPEIPPTTSEVSTPTT          838

Db  584 ---------TTPKEPEPTTPKEPEPTTPKEPEPTTPKKPEPTT                   617
                 | ||||||  ||||||||||||||  ||||
Qy  839 TKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTTKTPAATKPEMTTTAKDKTTER  898

Qy  899 DLRTPETTTAAPKMTKETATTIEKTTESKIATTIQVTSTTQDTTPFKITTLKTTTTLA    958
        |:|:|                                   |||| ||||:| |||
Db  618 ---TSPKI--------------------------------TTLKATTLA            631

Qy  959 PKVTTTTKKITITTEIMNKPEETAKPKDRATNSKAT-----TPKP-QKPTKAPKKPTS   1009
        |   | ||  |  |||| | ||  |  ||||||| ||     |||| ||||||||||||
Db  632 PKVTAPAE----EIQNKPEETTPASEDSDSKTTLKPEQKPTKAPKPTKKPTKKPTKKPTS  686

Qy 1010 TKKPKTMPRVRKPKTTPTPRKMTSTMPELNPSRIAEAMLQTTTRPNQTPNSKLVEVNPK  1069
        | | ||  | |||||||||||| ||||||||  ||||||||||||||||||| || ||
Db  687 TKKPKT-PKTRKPKTTPAPLKTTSATPELNTTP--LEVMLPTTTIPKQTPNPETAEVNPD  743

Qy 1070 SEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDG 1129
        | |  |||||||| | || |  |: | |  |  |||||| |||||||| | |||| |||
Db  744 HEDADGDGGEKP-LIPGPPVLPTTAIPGTDLLAGRLNRGININPMPSDETNLCNGKPVDG  802

Qy 1130 LTTLRNGTLVAFRGHYFWMLSPFSPSPSPARRITEVWGIPSPIDTVFTRCNCEGKTFFFKD 1189
        ||||||||||||||||||| |  |||    |||||||||||||||||||||||||||||
Db  803 LTILRNGTLVAFRGHYFWMLNPRPPSPPRRITEVWGIPSPIDTVFTRCNCEGKTFFFKD  862

Qy 1190 SQYWRFTNDIKDAGYPKPIFKGFGGLTCQIVAALSTARYKNWPESVYFFKRGGSIQQYIY 1249
        ||||||||| |:|||||||||||||||  |||:|||:|||| |||||||||||| ||| |
Db  863 SQYWRFTNDVVDPGYPKQIVKGFGGLTGKIVAALSIAKYKDRPESVYFFKRGGNIQQYTY  922

Qy 1250 KQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQYS-PARLAYQDKGV 1308
        ||||::||:|||||:||  |||  |||||||||:||:|||| ||:|| || | |||||
Db  923 KQEPMKKCTGRRPAINYSVYGEAAQVRRRFERAVGPFQTHTFRIHYSVPMRVSYQDKGF  982

Qy 1309 LHNEVKVSILWRGLPNVVTSAISLPNIRKPDGIYDIYAFSKDQYYNIDVPSRTARAITTRS 1368
        ||||||||| |||||||||||| |||||||||  | ||||||||||||||:||||||||||
Db  983 LHNEVKVSTMWRGPPNVVTSAIILPNIRKPDGVDYYAFSKDQYYNIDVPTRTARAITTRS 1042

Qy 1369 GQTLSKVWYNCP 1380
        ||||| |||||||
Db 1043 GQTLSKWVNCP  1054
```

```
Query Match             23.1%;  Score  1713.5;  DB 6;  Length 401;
Best Local Similarity   80.2%;  Pred. No. 2.7e-103;
Matches  320;  Conservative  24;  Mismatches  52;  Indels  3;  Gaps  3;

Qy  983 PKDRATNSKRANTTPKPQKPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRK-MTSTMPELNPT 1041
        || |||||: |||||||||| ||||||||||||| ||  | |||||||| | ||::||||||
Db    5 PKGRATNSQVTTPKPQKPIKAPKKPTSTKKPRT--PRVRKPKTTPTPPKTTTSAMPEPTPT   63

Qy 1042 SRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYL  1101
        | :||||||||||| ||||| :: ||:|| : | ||||:|||:: || |:|: | | | :
Db   64 S-LPEAMLQTTTRPTPTPNSEIIDVNSENEDGDAAEGEKPHMIFRPPVLTPIVIPGTEII   122

Qy 1102 PRVPNQGIIINPMLSDETNICNGRPVDGLITTLRNGTLVAFRGHYFWMLSPFSPSPSPARRI 1161
        | :|:|| |||| |||||| |||||||||||| |||||||||||||||| |  |:    |||
Db  123 VRGPSQGFGINPMFSDETNLCNGRPVDGLITTLIRNGTLVAFRGHYFWMLTPFTPPPPRRI  182

Qy 1162 TEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVA  1221
        ||||||||||||||||||||||||||| |||||||| ||:|||||||||||||||||||
Db  183 TEVWGIPSPIDTVFTRCNCEGKTFFFKGSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVA   242

Qy 1222 ALSTARYKNWPESVYFFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFE  1281
        ||| |:||  ||||||||||| ||| ||||||:||: ||||:|| :||| |||||||||
Db  243 ALSIAQYKSRPESVYFFKRGGSVQQYTYKQEPTQKCTGRRPAINYSVYGETAQVRRRFE   302

Qy 1282 RAIGPSQTHTIRIQYSPARLAYQDKGVLINEVKVSILWRGLPNVVTSAISLPNIRKPDGY  1341
        ||||||| ||:|| |: |:: ||||| |:|||||| |||||||||||||||||||||||
Db  303 RAIGPSQVHHIRIHYTPVRVPYQDKGFLHNEVKVSTLWRGLPNVVTSAISLPNIRKPDGY  362

Qy 1342 DYYAFSKDQYYNIDVPSRTARAITTRSGQTLSKVWYNCP 1380
        |||| |||||||||||||||||||||||||||: ||||||
Db  363 DYYALSKDQYYNIDVPSRTARAITTRSGQTLSNIWYNCP  401
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Cotton M.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   1079 AA;  110532 MW;  8DBDE3824CF80CA1 CRC64;

Query Match             13.3%;
Best Local Similarity   29.6%;
Matches  332; Conservative  131; Mismatches  481; Indels  177; Gaps  44;

Score 982.5; DB 5; Length 1079;
Pred. No. 1.4e-55;

Qy    171 QKKLKVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKV--------------TTPDTSTTQ-HNKV 222
              :|| :||  |:   :||
Db     27 QKELEIADCRANQTPHTMLP-----------------STLTSVDMETPSTLVLSSTPISSSTPIKETT 77

Qy    223 STSPKITTAKPINPRPSLPPNSDDISKETSLTVNKETT--VETKETTTTNKQTSTDGKEKTT 281
              |||  |||                   ||| |     |||  |||   ||   |
Db     78 TTIAPETTSTEP------PSSSTTPVQTTTTTAPETTSTEAPSSSSTTPVQTT----TT 124

Qy    282 SAKETQSIEKTSAKDLAPTSKVLAKPTPKAETT-TKGP-ALTTPKEPTPTTPKEPASTTP 339
              |   :|   : |||       |   ||| ::||  |   |||  |||
Db    125 TAPETTSTEPPSSS-----TSPVQTTTTTAPETTSTEAPSSSTTPVQTTTTTT 180

Qy    340 KEPTPTTIIKSAPTTPKEPAPTTTKSAPTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 399
              |  |||   |||   |||    |||   |||    |||   |||
Db    181 PSSSTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPE---TTISTEP-PSSSTTPVQT 236

Qy    400 TTIKSAPTTPKEPAPTTPKKPAPTTPKEPEPTTTPKEPAPTTTKEPAPTTPKEPA 459
              ||   |||   |||   |||   |||   |||   |||   |||
Db    237 TTTTAPETTSTESPSSSSTTPVQTTTTTTAPETTSTEP-PSSSTTPVQTTTTTAPETTSTEP 295

Qy    460 PTAPKKPAPTTPKEPAPTTPKEP---------------APTTTKEPSPTTPKEPAPTTPKS 505
              |:   |||   |||   |||                 |||   |||   |||
Db    296 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTT 355

Qy    506 APTTTK--EPAPTTIKSAPTTPKEPSPTTTKEP-APTTPKEPAPTTPKKPAPTTPKEPAP 562
              |||     |||   |||   |||   |||  |||   |||   |||   |||
Db    356 APETTTREEPPSSSTTPVQNTTTTAPETTSTEPPSSSTTPVQTTTTTT---APETTTSTEPP 411

Qy    563 TTPKEPAPTTTKKPAPTAPKEP---APTTPKETAPTTPKLTPTTPEKLAPTTPEKPAPTT 620
              |||   |||                 |||   |||    |   |||
Db    412 SSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTT----PSSSTTPVQTT 467

Qy    621 PEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEP---APTTPKEPAPTTPKEPAPTTPK 678
              |   |||   |||   |||   |||         |||   |||   |||   |||
Db    468 TTTAPETTSTEP-PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTES 525

Qy    679 ETAPTTPKGTAPTTPLKEPAPTTPKKPAPKELAPTTPKEPTSTTSDKP-APTTPKGTAPTT 737
              |   |||   |||   |||   |||   |||   |||   |||
Db    526 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTT 585

Qy    738 PKEPAPTTPKEP--APTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKP- 794
              |||   |||     |||   |||   |||   |||   |||   |||
Db    586 APE---TTTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTSDKP- 642

Qy    795 APTTPKETAPTTPKEPAPTTPKK----PAPTTEPPPPTTS---------EVSPTTTTKE 841
              |||   |||   |||   |||       |||
Db    643 SSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSSNTPVQTTTTTAP 702

Qy    842 PTTIHKSPDESTPELSAEPT--PKALENSPKEPGVPTTKTPAATKPEMTTTAKDTTTERD 899
              |||   |||   |||         |||   |||   |||   |||
Db    703 ETTTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTP 762

Qy    900 LRTT----PETTTAAPKMTKETATTTTEKTTESKTIATTTQVTSTTTQDTTPKTTTLKTT 955
              |   |||   |||   |||   |||   |||   |||   |||
Db    763 VQTTTTTAPETTSTEPPSSSNTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETT 820

Qy    956 TLAPKVTTTKKTTITTTEIMNKPEETA-KPKDRATNSKATTPKPQKPTRAPKKPTSTKKPK 1014
              |   |||   |||   |||   |||   |||   |||   |||
Db    821 STEPPSSSNTPVQTTT--TTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPV 878

Qy   1015 TMPRVRKPKTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTTPNSKLVEVNPKSEDAG 1074
              |   |||   |||   |||   |||   |||   |||
Db    879 QTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPET---TSTEP---PSSSTTPVQTTTTTAP 932

Qy   1075 GAEG-ETPHMLLRPHVFMPEVTP---DMDYLPRV-------PNQIIINPMLSDETN 1120
              |   |||   |||   |||   |||
Db    933 ETTTSTEPPSSSTTPVQTTTTTVPCDCSLSYIDRVVYPFTTEWWENKRDITIIQSYDSPRRT 992

Qy   1121 ---------ICNGKPVDGLLTTLRNGTLVAFRGHYFWMLS--------PFSPPS 1156
              |   |||   |||   |||
Db    993 AFSFVSFTDIGTCTAILICIYTSEGGISNL-NATL--------FIGLSDGSSIDLPF---- 1040

Qy   1157 PARRITEVWGIPSPIDIVFTRCNCEGKTFFFRKDSQYWRFTN 1197
              |   |||
Db   1041 --------YNVPIGLE-IWPEINCEGNK------WSYNN 1064
```

```
RESULT 6
Q917S1

ID   Q917S1      PRELIMINARY;      PRT;   1049 AA.
AC   Q917S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003495; AAG22353.1; -.
DR   FlyBase; FBgn0030561; CG5228.
SQ   SEQUENCE   1049 AA;  954DD629E7619671 CRC64;

Query Match             12.8%;  Score 946.5;  DB 5;  Length 1049;
Best Local Similarity   34.4%;  Pred. No. 2.9e-53;
Matches  376;  Conservative   69;  Mismatches  480;  Indels  169;  Gaps   56;

Qy    82 EVHNPTSPPSSKKAPPPSGASQTIKSTTKRSPKPPPNKKKTKKVIESEEITEEHSVSENQE 141
            : :  :       ::                                    ::
Db    33 EIGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPT------EGTTAKPTTLKPTE 86

Qy   142 SSSSSSSSSTIWKIKSSKNSAANRELQKKLKVDNKKNRTKKKPTPKPVVDEAGSG 201
            : :
Db    87 GTTAKPTT-------LKPTEGTSA-----KPTTLKPTE--GTTAKPTTLKPT---EGTS 128

Qy   202 LDNGDFKVTTPDIS--TTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETT 259
            | |   |  |       |       |       |             |       |
Db   129 AKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKP---TTLKPTEGTSAKPTT 185

Qy   260 VEFKETTTTTNKQT--STDGKEKTTSAKET---QSIEKTSAK--DLAPISKVLAKPT----PK 310
            |     |             |   |    |        |            |      |
Db   186 LKPTEGTSAKPTTLKPTEG----TSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTT 241

Qy   311 AETTTTKGPALTTPKEPTPTTPKPEPTTPKEP---PTTIKSAPTTPKEPA 364
            | ||    |  |   |   |     |             |   | |   |
Db   242 EGTTAK-----PTTLKPTEGTTAKPTTLNPTEGTSAKPTTLKPTEGTTAKPTTLNPTEGTS 297

Qy   365 APTTPKEPAPTTPKEPA----PTTPKEPAPTTTPKEPAPTTPKEPAPTTPKEPAPTTKKPAP 421
            | || |   | |                | ||       | |      | |
Db   298 AKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAK----PTTLK---PTEGTSAKP 352

Qy   422 TTPKEPAPTTPKEPT--PTTPKEPAPTTPKEPAPTTPKEPA----PTAPKPAPTTPKEPAP 476
            | |   | ||       | |      | |    | |       |  | | || |
Db   353 TTLKPTDGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG 412

Qy   477 TTPKEPAPTTKEPSPTTPKEPAPTTPKEPAPTTPKEPA---PTTTKSAPTTPKEPSPTTT 533
            | |       |   | |      | |    | |         |   | || |
Db   413 TTAK----PTTLKPTEGTTAK---PTTLKPTEGTTAK---PTTLKPTEGTSAKPTTLKPTEGTT 466

Qy   534 KERAPTTPKE---PAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKRPAPTTAPKE---PAPT 587
             | | || |      | |    | |      | |      |    | |          |
Db   467 AKPTTLKPTEGTTAKPTTLKPTKGTSAKPTTLK-PTEGTTAKPTTLKPTEGTTAKPT 525

Qy   588 TPKETAPTT---PKKLTPTTPEKLAPTT--PEKPAPTTPEELAPTT--TPEBPTTPTPEEPA 641
            | |             | |         |  | | |       |      | |
Db   526 TLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGT 585

Qy   642 PTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKPEPAPTTP 701
            | |         |   | ||    | |      | |      | |     || |
Db   586 TAKPTTLKPTEGTSAQPTTLK---PTEGTSAQPTTLKPTEGTSTAK---PTEG 636

Qy   702 KKPAPKELAPT-----TIKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTP 754
             | | |           |             | |     | |      | |
Db   637 TSAQPTTLKPTEGTTAKPTTLKPTEGTSAK---PTTLKPTDGTTA 691

Qy   755 KGTA--------PTTLKEPAPTTPK---KPAPKELAPTTKGPTSTTSDKPAPTTPK 800
             | |        | ||                | | |     |           | |
Db   692 KPTTLNPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPT 751

Qy   801 ETAPTTPKEPAPTT--PKKPAPTTPETPPPTSVSTPTTTK------EPTTHKSPDE 851
            |  | || |                     |    | |           |
Db   752 EGTRTTAK---PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAQPTTL--KPTE 806

Qy   852 STPELSAEPTP-KALENSPKEPGV--PTTKT---PAATKPEMTTTAKDKTTERDLRTTPE 905
            :  : | |       :  | |       | |      | |                |
Db   807 RT---SAQPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAQPTTLKPTE---LKPT-E 858

Qy   906 TTTAAPKMTKETATTTTEKTTESKITATTTQVISTTTODTTPFKITTLKTTTIAPKVTTTIK 965
            | |        |              |      |          |  | |          | |
Db   859 GTSAKPTTLKPTEGTTAKPT-----TLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTE 914

Qy   966 K---TITTTEIMNKPEETAKPKDRATNSKATTTPKPQKPTKAPK---KPT--STKKPKTMPR 1018
             |              |   | |              | |         | |        | |
Db   915 KPTTLKPTEGTSAKPTTLKPTE--GTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKP 973

Qy  1019 VRKPKPTTPTI---PRKMTSTMP-ELNPTSRIAEAMLQTTTTRRNQTPNSKLIVEVNPKSEDAG 1074
            : | | ||         | |      |       |      |               :
Db   974 TEGPSAKPTTLKPTERTSAQPTTLKPTERTSAQ--PTTLKPTEGTTAKPTTLKPTKGTSG 1031

Qy  1075 GAEG-ETPHMLLRP 1087
            | |
Db  1032 QANNFETKRKERRP 1045
```

```
RESULT 7
O76894

ID   O76894           PRELIMINARY;      PRT;   1795 AA.
AC   O76894;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:56G7.1 PROTEIN.
GN   EG:56G7.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
```

```
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003421; AAF45644.1; -.
DR   EMBL; AL031028; CAA19845.2; -.
DR   FlyBase; FBgn0025390; EG:56G7.1.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF01607; Chitin_bind_2; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00494; ChtBD2; 2.
SQ   SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match          12.2%;  Score 902; DB 5; Length 1795;
Best Local Similarity 25.9%;  Pred. No. 3.8e-50;
Matches 330;  Conservative 163;  Mismatches 537;  Indels 246;  Gaps 43;

Qy    1 QDLSSCAGRC-------------------GEG-YSRDATCNCDYNC-----QHYMECCP--- 34
          :  ::: ||||                          ::: :
Db  124 ESFNASSGRCVQHVPQHRPDHRPQCQKEGRFPHPHDCKYYRCDKNRTQPWLFPACPAGT 183

Qy   35 ---DFKRVC-------------------TAELSCKG------------RCFE--SFERGRECDCDAQCKKYD 70
            : :||                      :|  ||             :|:|     :: |: |  ::| |
Db  184 IFSPVERKCLPGDQCPSTISDSGSYIPQNCELKFPECAEEGTFRSPTDCALYYTCRLQE 243

Qy   71 K------------CCPDYESFCAE-----------------VHNPTSP------PSSKKAPP-- 97
                         :| :  | |                     : |||          ::| ||
Db  244 SGTYLQTRFKCPGSNSFDLERKLCRPRSEVDCFDFVPGPVQVPYAPQPYYPPYPPAAPPLY 303

Qy   98 -----PSGASQ---TIKSTTKRSPKPPNKKKTKVIESEEIEEHSVSENQESSSSSSSSSS 149
             :| |:|       ::  :: |                                ::
Db  304 EEDDYDTGAREQOPALKSEKLQVAAEGFEKPSLNVVVLQTTLEPSTAYHKYPAYPSYPS 363

Qy  150 SSSSTIWKIKSSKNSAANRELQK------KLKVKDN----------- 179
           ::: |  |  :  ::  || |       :|| :|
Db  364 YEYSS--HHRGKERAAENLELEKEGVPRKLKLSENIVIQOPEPATAANTREPLNDINKYQ 421

Qy  180 --------KKNRKRKKPTPKPPVVDEAGSGLDNGDRKVTTPDTSTTQHNKVSTSPKITTA 231
                    ||:|:|  | |  |          |:|  :     :| :|  : :||: |
Db  422 YKRYTYGTDKNDVTEAPEIKSPL----KGLHLSENIVILPETTTT--TTTTTKPVVLTC 474

Qy  232 KPINPRPSLPPNSDT------------SKETSLTVNKETTVETKETTTNKQTSTDGKEKT 280
            | :|| :   :            :||::| |:|:|: | :|||:|: |:||  |:||
Db  475 PTISPPDTTPKPSTTTAVTKSTPKISSTEQHSTTTAKTTTTKRPTTVTTEKTISSATEKPT 534

Qy  281 TSAKETQSIEKTTSAKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPPTTPKEPASTTPK 340
          | ::|| | :||: ||  | |: ||   |:|  ||  : | |||||| |||||| |||||
Db  535 TVTTTTTTQKRSTTTHNTSPDTKTTIRSTTLSPKTTTPSTTTPSTTTPST------TTPS 588

Qy  341 EPPPTTIKSAPTTPKEPAPTTPKEPAPTTTK-EPAPTTPKEPAPTTTKEPAP 399
          ||| :: :|||||| ||||||  | |:|||   ||| |||:|| ||||||
Db  589 TTTPST----TTPSTTTPSTTTTVKSTHRPTTISQKTTTASTTTKK----TTTSPKTT 639

Qy  400 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTKKEPAPTTKKEPAPTTPKEPAPTTPKEPA 459
          ::| |||||| |||||| ||||||:|||||||:|||||| :|||||||||||||||||||
Db  640 KTTDIPTSTTSKLLSTTTQKTTTTHKFTAATTSTEPKFTTIVSTTTKKSTESSSPK 699

Qy  460 PTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKEPAPTTK 519
          |:  || ||||||||||||||||||||:|||:|||||||||||:|||||||
Db  700 PTSSTGKPTTTPKPSTRTTPTTTPKVTTTTQITTTPLRSSTETTTSTQPPPTTTTQPTTTT 759

Qy  520 SAPTTPKEPSPTTTKEPAPTTPKEPAPTTPK-------KPAPTTPKEPAP----- 562
           ||||||||:|||:|||||||||||||||||        ||||||||||
Db  760 TLTVTPKTSTTTTTTEKPITSSPKPTTTTSTAPNTTKVAITTQKETTPTQSTSTTI 819

Qy  563 ---TTPKEPAPTTIKKPAPTTAPKEPAPTTPK-ETAPTTPKKLTPTTTPEKLAPTTPEKP- 616
             |||||||||::||||||| ||||||||||  ||||||||: :|||||  |||||| |
Db  820 FTTKRTTTTNNPEPTSTEKPITSTTTPKPSTTTPKTSTVASSTEKTTISSPKPTTEKSTENPT 879

Qy  617 -----APTTPEELAPTTPEEPTPTTPTTPEEPAPTTPKAAAPNTKPEPAPTTPKEPAPTT- 667
               |||||::||||||::| |||| ||::|||||||  |||:|||||||||||||||
Db  880 TNSVKTSALTSSTQRATSTTTSEPTKTT-QNITTTTPKPTTLKTSTQEATTSTQKVSTVTI 938

Qy  668 TPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPT--STTSDKP 725
          |||||||||||| |||||||||||| |||||||||||||||||||||||   :|| |||
Db  939 TTKKATTESSPLTTLSTEEPNTTPKPLRTTTPTTTPSVTATTRITTTTTISESSTETTTSTQKP 998

Qy  726 APTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP---TTTK 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||||
Db  999 KSTTPTSTTRTTPKVTTTVIVSTQNPTTTTTSKTSTVTI-----TTP-NPSPSTQRPTTTTR 1052

Qy  785 GPTTSTTSDKPAPTTPKETAPTTPKEPAPTTTKKKPAPTTPTETPPPTTSEVSTPTT-TKE-- 841
           |||||| ||||||||| |||||||||||| ||||||| |:| ||| ||||:||||  ||
Db 1053 QPTSTTASTTTSIGTTTRIPTTTNPQNSTSSTTDLTTVTTRPPCPDPDDSTISDKNTNTACTQELQ 1112

Qy  842 --------PTTIHKSPDESTTPELSAEPTTPKALENSPKEPGVPTTTKTPAATK 884
                  |||:|| | :||||| ||||||| |::|| ||  |||||||| |:|
Db  885 PEMTTTAAKIDKTTER---DLRTTTPETTTAAPKMTKETATTTEKTTTESKITTATTTTQVTISTTT 941

Db 1113 QVNLLELQSPQKQEQFTHRTHTHFALTGSRNTLGGQOEVPDYMDDAPSSAAEAESGOATAKA 1172

Db 1173 PTMSTLAAAHLLQKLFHIISTTPSREHAP--TQRPSSQOPSSSQORSR-GVTIAQMARHNL 1229
Qy  942 QDTTPFKITTLKTT------TLAPKVTTIKKTIITTTEIMNKPEETAKPKDRATNSKATTPKP 997
          | || || ||| |       |||  |||:||| : || ||:|||| | |:|  | |||||
Db 1230 ATSKPFIAHSLRLSIQQLASTQKRSIPKTLVTHHNTTKEPED-SEYYDSETSEQYDEDN 1288

Qy  998 QKPTKA-PKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1056
          :|| || ||||:|:| |:|  | |||:|   |:|:|:|  ||||||: || | |||:|
Db 1289 EVLDKTQPRAMSSTTVAAVLPAVPSTTTEREPQK-TSSSP--SPT----KATSSTTTQPI 1341

Qy 1057 QTPNSKLVEVNPKSED 1072
          ::|||| |: :| :|
Db 1342 EITTTGDLEYDSGSGSSD 1357
```

RESULT 8
O96449
O96449;

ID   O96449   PRELIMINARY; PRT; 1489 AA.
AC   O96449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; Stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goernhardt B.;
RC   STRAIN=RACE 1-11;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061185; AAC72308.1; -.
SQ   SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

```
Query Match           11.7%;   Score 863.5;  DB 10;  Length 1489;
Best Local Similarity 31.7%;   Pred. No. 9.6e-48;
Matches 368;  Conservative 55;  Mismatches 509;  Indels 229;  Gaps 51;

Qy   121 TKKVIESEEITEEHSVSENQESSSSSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVRDNK 180
            ||||||||||||||||:|  |||||||||||||||||||||||    ||| ||:|
Db   282 TPYVGYSTEETEGQHVTGGYEPSDETEAPTEGTTY--VPREETTAAPSE----DTTYAP 334

Qy   181 KNRTKKPTPKPVVDEAGSGLDNGDRFKVTFPDTTSTT----QHNKVSTSPKITTAKPI 234
            |||||||||| ||||||||| ||| |||:|||||||       ::| | |||
Db   335 REVTPYAPTEKPVDVEETTYVTEESTYAPTKSETNAPTERMHYAHIEKPCCDEVTMYAPT 394

Qy   235 NPPRSLP------------              ---PNSDT---SKETSLITVNKETTVETKETTT---TNKQT 272
                                |:|           :|:| |||| |||||||:||||| --- |||:|
Db   395 EETTYAPTEETT-YAPTEEMYAPIEETTYAPTEETTIYPTEETTIYAPTEETTIYAPTEKTI 454

Qy   273 STDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALITTPKEPTPTTPK 332
            |||   |:  | ||||||||||||||| |||||| ||||||||||||:|||||||||
Db   395 EETTYAPTEETTYAPTEEETTIYAPTEETTTYPTEETTIYAPTEETTIYAPTEETTYAPT 454

Qy   333 EPASTTPKEPTPTTIIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT---TTKEPAPTTPKEPA 389
            |||||||| || |::| ||:| ||||||    |||:| ||||||     ||||||||||
Db   512 APAEETPYEPTEEPT-TYAPTEEETTIYAPTEETTIYAPTEETTIYAPAEEETPYEPT 570

Qy   440 PKEP-----------                        --APTAPKKPAPTTPKEPAP 476
              ||||                                  ||| |||||||||||||
Db   571 EETTYAPTEETT-YAPTEEETMYAPIEETTYAPAEEETPYEPTEETTIYAPTEETT 629

Qy   390 PTTTKEPAPTTTKSAPT------              --TPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTT 439
            |||||||||||:|||        || ||||:|||| |||||||||||||||
Db   571 EETTYAPTEETT-YAPTEEETMYAPIEETTYAPAEEETPYEPTEETTIYAPTEETT 629

Qy   477 TTPKEPAPT---                          --TTKEPSPTTPKEP-----              --APT-TTKSAPT---                  --TTKEP---- 513
            ||||||||                              ||| ||:| ||||||
Db   630 TYAPTEETTYASTEEETTIYAPTEEETTIYAPAEEETTIYAPTEETTIYAPTEETTYAP 689

Qy   514 --APT--TTKSAPT---                      --TPKEPSPTTPKEP-----              --APTTPKEPAPTTPKKPAPTTPKEPAP 562
              ||| |||:|                            ||| ||:|||||||||||||
Db   690 TEETTYAPTEETTYAPAEEETPYEPTEETTIYAPTEETTIYAPTEETTYGPTEET 749

Qy   563 T---                                  --TPKEPAPTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPKKPAPTT 620
            |                                    |:||||||||||||||||||
Db   750 TYAPTEATTYAPTEETPYAPTEETTYEPTGETTIYAPTEETTIYAPTEEETTYAP 809

Qy   621 PEELAPT---                            --APT---                      --TPKAAAPNTPKEPAPTTPK 662
            |||                                  ||:          || |||
Db   810 TEETPYEPTEETTIYAPTEEPYEPTEET---TYTPTEETTIYAPTEKTIYAPTE 867

Qy   663 EP-----              --APTTPKEPAPTTPKETAPTTPKGTAPT---               --TLKEPAPTTPKKPAPK-ELAPT- 712
              ||||||||||||||  ||:| || ||||||||||||| |||||
Db   868 EETTYAPTEETTPYEPTEETTIYASTEEETTIYAPTEETTYAPAEEETPY 927

Qy   713 -TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP-----      --APTTPKGTAPT---- 760
             |:| || ||:| |||||||||||||||:|||   |||||
Db   928 EPTEETTYAPTEETTYAPAEEETPYEPTEETTIYAPTEETTYAPTE 987

Qy   761 -----                    --TLKEPAPTTPKKP-----      --APKE---         --LAPT----          --TTKGPTSTTSDKPAP 796
                                      |||||||||        ||:         ||||        ||||||||:|
Db   988 EETTYAPTEEETMYAPIEETTYAPTEETTIYAPTEETTY 1047

Qy   797 TTPKETAPTTPKEPAPTTPKEPTPPPTTSEVSTPTTTKEPT--TIHKSPDESTP 854
            |||||||||||||||:||  :||||||| |||||||||| -- ||:|:|:||
Db  1048 ASTEETTYAPTEEETTYAPAEEETPYEPTEETTIYAPTEETTIYAPTEETT 1107

Qy   855 ELSAEPTPKA-LENSPKEPG-----                --VPTTKTPAATKPEMTTTAKDKTTERDLRTTP---- 904
            ||:|||||  |:|:                          ||| |:|| ||| |||:
Db  1108 EETTYAPAEEETPYEPTEETTIYAPTEETTYGPTEETTY-APTEATT 1166

Qy   905 ETTTAAPRKMTKETATTIEKTTESKITATTTQVTSTTTQDTTT-----           --PFKITT---     951
            ||||  |:|   | |||  | ::|  ||| ||:|:||              ||:|:
Db  1167 YAPTEETTYAPTEETTIYGPTGETTIYAPTEETTIYAPTEETTIYAPTEEMYAPIEETTYPEPT 1226

Qy   952 LKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAT--NSKATTPKP-QKPTKAPKKPT 1008
            ||||| |||:|||||||||||:|||||||||        |:||:|| |||:|||||
Db  1227 EETTYAP-----              --TEETTIYEPTEETTIYAPTEETTIYAPTEETMYAPIDEETTYGP 1281
```

```
RESULT  9

Q20007
ID      Q20007;                          PRELIMINARY;          PRT;      1274 AA.
AC      Q20007;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT      01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE      COSMID F35A5.
GN      F35A5.1.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RX      MEDLINE=94150718; PubMed=7906398;
RA      Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA      Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA      Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA      Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA      Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA      Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA      Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA      Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA      Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA      Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans.";
RL      Nature 368:32-38(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Leimbach D.;
RL      Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Waterston R.;
RL      Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Waterston R.;
RL      Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; U46675; AA52641.1; -.
RL      EMBL; U46675; AA52641.1; -.
SQ      SEQUENCE   1274 AA;  138065 MW;  07F6B0D292C1799F CRC64;
```

```
Db  1282 TEETTYAPTEEATTYAPTEETTPYAPTEETTIYEPTIGETTIYAPTEETTYAPMEET 1341
            || |:|       :| ||||||||||||||||||||:|  ||| |:|| |
Qy  1009 -----          --STKKP-KTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP 1055

Db  1342 PYEPAEESTSTVSYTEKPCNTEEFTDEFTDEPTIDE--PSDEPTDEPTDEPTIDLPTDEPSTP 1399
            |:|  |||| || :| || |||||:||||:||:||   ||:||||||||| |||:| |
Qy  1056 --NQTPNSKLVEVNPKSEDAG 1074

Db  1400 CDNQGINGIGVENKVRYNNAG 1420
```

```
Query Match           11.4%;   Score 846;  DB 5;  Length 1274;
Best Local Similarity 27.6%;   Pred. No. 1.1e-46;
Matches 313;  Conservative 119;  Mismatches 455;  Indels 248;  Gaps 60;

Qy    90 PSSKKAPPSGASQTIKSTTKRSP-----               --KPP---NKKKTKKVIESEEITEEHSVSENQES 142
            |||||||||| |:|||||||     |||         ||:|:||||||||||||||||||
Db   185 PSPKKAAPSKEHDPIVPPTIKNPAKKWKPWEDDEVPTEIKKEPEPATRKVPALKKKEP 244

Qy   143 SSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK-KNRTKK-KP-----     --TP---- 190
            |:||||||||||||||||||||||||||||||||||| |||||| ||     |
Db   245 STSVKPVSDPSPYKKV------                 --PVKKEEVPPPPIKNPTKKWKPPWEDETPVEEVK 294
```

```
RESULT 10
Q9SPM0
ID   Q9SPM0
AC   Q9SPM0;
```

```
Query Match             11.0%;  Score 816;  DB 10;  Length 1315;
Best Local Similarity   23.5%;  Pred. No. 9.9e-45;
Matches  265;  Conservative  106;  Mismatches  486;  Indels  270;  Gaps  35;
```

Alignment (right column block):

```
Qy   49  RCFESFERGRCECDCDAQCKYDKCCPDYESFCCAEVHNPTPSPPSSK-KAPPPSGASQTIKS  107
         :|
Db  390  QCAPVIARPVEVDCS------------------KHVCAGYPTPGGPPSSSVPGKPPSVPGK----  432

Qy  108  TTKRSPKPPNKKTKKVVIESEEITEEHSVSENQESSSSSSSSSSSSSTWKIKSSKNSAAN  167
         :|
Db  433  --PAAPAPMPTHHPPPDVSPEPLPEPSPV------------------PAPAPMRMPTLRSEPPAD  476

Qy  168  RELQKKLKVKDNKKNRTKKPPTPKPPPVVDEAGSGLDNGDFKVTTPDYSTTQHNKVSTSPK  227
         |:|
Db  477  EYI-------------------PTPPVPAKSPPG-----TSPPASRGA-PPLQAQPP  508

Qy  228  ITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTINKQTSTDGKEKTTSAKETQ  287
         :|:||
Db  509  AASSPPATPVKSSPPPAAV------------------------------------  527

Qy  288  SIEKTSARDLAPTSKVLAKPTKAETTTKGPALTTPKEPTPPTTPKEPASTTPKEPPTTI  347
         :|:||
Db  528  -VLPPPAKTPSPEAPV-ASPPEEAPVSSPQPOVKSPPPAVSSPPPMKSP--PPPARV  583

Qy  348  KSAPTTPKEPAPTTPKSAPTTP-KEPAP--------TTTKEPAPTTPKEPAPTTPKEPAPT  399
         |:||
Db  584  ASPPPLMKSPPPAPVASPPQPLKSPPPPVLWLSTPSVKSPPPPVVASPPPPVKSPPPL  643

Qy  400  TTKSAPTTPKE----PAP---TTP----KKPAPTTPKEPAPTTPKE-PTPTTPKEPAPTTK  449
         |:||
Db  644  APVSSPPPPVKLPPLPAPGKSTPPEEEKPTPPTPVKSSPPPEEPVSSPQARKSSSP------  703

Qy  449  EPAPTTPKEPAPTARKKPAPTTPKEPAPTTPKE-PTPTTPKEPAPTTKSAPTT  509
         |:||
Db  704  PTPPSTPSKPPPPSVKLPPLPAPGKSTPPEEEKPTPPTPVKSSPPPEEPVSSPPQARKSSSP  756

Qy  509  TTKEPAPTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKEPAPTTPKSSPPLAPVSS  569
         |:||
Db  757  TTKEPAPTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKEPAPTTPKSSPPLAPVSS  816

Qy  569  APTTTKKPAPTAPKEPAPTTPKKLTPTTPEEKLAPTTPKEPAPTTPEELAPTT  628
         |:||
Db  817  PPQVEKTSPPPAPVSSPPPFPKSSPPLAPVSSPPQVEKTSPPPAPVSSPPLEPKPSSPPS  876
```

Alignment (left / continuation block):

```
Qy  191  KPPVVDEAGSGLDNGD----FKVTTPDYSTTQHNKVSTSPKITTAKPI-NPRPSLPPNSD  245
         |||  ||  :|
Db  295  EPPVPEKKAPVLKKKDPAPAAKARDPSPSKAAPKKVEPSSPVVPPTPVKKNPVKKPWE  354

Qy  246  TSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLA  305
         :|                |||
Db  355  VDDEPAEEVKKPSAPEKKTPVLKKEPEPSSTTPSSDPSPKKAAPAVKPRDSSPKKATPL  414

Qy  306  KPTPKAE------------TTTKGPA------------------------KEPASTTPK  340
         |||                |||||                         |:||||||
Db  415  QADPKAQEVPPTPVKNPVKKYKPPWEVDDEDPVEEVKQPEAPAKKTPVLKRKEPAAKDTA  474

Qy  341  EP-----TP-TTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTTKEPA-----------  381
         :|       :|
Db  475  KPATSKTPETPEKKDPVKPRDSSPKKVAAKPDSAQAPA-TPVKNPVKKWRPPWEDDETPA  533

Qy  382  -----PTTPKEPAPTTPKEPAPT-----APTTPKKPAPTTP  424
              |||||||| |||
Db  534  DDVSKPIDAKTPSLAKKDPAPAKESLKPADTKAPAKPRDPSPKKVAPTAPEKKTPVLA  593

Qy  425  KE------PAPTTPKEPTPTTPKEPAPTTPKEPAPTTPK----EPAPTARKKPAP----  473
         |:
Db  594  KKEPAGPADSKTKEPEKSKPRDPSPKKAVPAPAKPVKTEVAPAAVKKPEPISKPKDTAPKK  653

Qy  474  RAPTTPKEPAPTTTKEP-----SPTTPKEPAPTTPKEPAPTTTKEPAPTTTKKPAPTTK  528
         |:||
Db  654  AEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPEPEKK-TPVLAKKAPIKPDSEA  711

Qy  529  SPTTTKEPAPTTPK---EPAPTTPKKPAP--TTPKEPAPTTPKEPAPTTTKKPAPTA-PKE  583
         |:||
Db  712  AADPVSGSPSKDPKLAKKAAPVKRDPSPMKAVPIKPAPKT--EVPPAVVKKPEPVAKSRD  769

Qy  584  PAPTTPKETAPTTPKKLTPTTP------EKLAPTTPKRPAPTTPEELAPTTPEEP----T  633
         |:||
Db  770  PSPKKAK-AEPNSP--VVPPTPVKNPVKKWKPPWEDDDAPAEPAVNVPEPEKKTPVLAKKT  826

Qy  634  PTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTP--PKEPAPTTPKETAPTTPKGTAPTT  692
         |:||
Db  827  PVKPRDPSPKKAVPAKPSTKDAPPVSVKKPEPVSKPEPSPKKAAEPNSPVVP-----PTP  882

Qy  693  LKEPAPTTPKPAPKELAPT-TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAP  751
         |:
Db  883  VKNPVKKW-KPPWEDDDEPTEEVKKPSE--PEKKTPVLAK-KEPEKPRD-APKVAAKPRD  937

Qy  752  TTPKGTAPTTLKEPAPT---TPKKPAP------KELAPTTTKGPDTS-----TTSDK  793
         |:||
Db  938  PSPRKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEAPPTPVKNPVKKWKPPWEDDDE  995

Qy  794  P-------APTTPKET----APTTPKEDPTTPKKRPAPTTPETPPPTTPSEVSPTTTKEP  842
         |:||
Db  996  PAEPVSAPEPEKKTPVLAKKAPAKPRDP--SPKKAAPVAAK-PDPKIPEV-PPTPVKNP  1050

Qy  843  TTIHKSP------DESTPELSA-EP------TPKALENSPKEPGV------PTTKTPA  881
         |:||
Db 1051  VKKWKPPWEDDDEDDEPSEPVSAPEPEKKTPVLAKKAPTKPATKPDSEAAADPVSGPTSKDPK  1110

Qy  882  AT---------KPEMTTAKDKTTERDLRTTPETT-TTAAPKMTKETATTTEKTTESKITPATT  933
         |:
Db 1111  LSKKAPVEKPKPTTDPKDDKLKPSPAAKPEKRAPAEPAAPKKWKPVWEDDDPDPEADFTVPA  1170

Qy  934  TQVTSTTTQDDTTPFKITFLKTTLAPKVTTEIMNKPEETTAPKDRATNSKAT  993
         |:
Db 1171  PSKKPDTEDPADPLG--------GPKTKDPK--------LNKKAPAEKPTEK  1206

Qy  994  TPKPQKPFKAPKKPTSTKKPKTMPRVRKPK------------------  1036
         |:
Db 1207  -PKPKEVSKEPEKPTEPPKP-AAPKKWKPPWEDDDEDPEPEADFTMPAPKKPDTEDP  1259
```

```
RESULT 11

ID   P70670      PRELIMINARY;    PRT;   2187 AA.
AC   P70670;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE   MUSCLE-SPECIFIC FORM GP220).
GN   NACA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96312450; PubMed=8698236;
RA   Yotov W.V., St-Arnaud R.;
RT   "Differential splicing-in of a proline-rich exon converts alphaNAC
RT   into a muscle-specific transcription factor.";
RL   Genes Dev. 10:1763-1772(1996).
DR   EMBL; U48364; AAB18734.1; -.
DR   EMBL; U48363; AAB18732.1; -.
DR   MGD; MGI:106095; Naca.
DR   InterPro; IPR002715; NAC.
DR   InterPro; IPR003037; TS-N.
DR   Pfam; PF01849; NAC; 1.
DR   Pfam; PF02094; TS-N; 1.
SQ   SEQUENCE  2187 AA;  220599 MW;  003646AA864DEBFD CRC64;
```

```
Query Match             11.0%;
Best Local Similarity   27.4%;
Matches  311;  Conservative  128;  Mismatches  449;  Indels  246;  Gaps  51;

Score 812;  DB 11;  Length 2187;
Pred. No. 3e-44;
```

```
Qy    629 PEEPTPTPTEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGT 688
Db    877 SVSSPPPTTVKSSPPPAPLSSPPMTPKSSSPPEAEKSSPPLAPISSPPSEPKSP 936

Qy    689 A------------PTTLKEPAPTTPKPAPKELAPTTKEPTSTTSDKPAPTT----TPKG 732
Db    937 SSSPPMVEKTSPPPATVSSPPPPPKKSSPPP---APVSSPPPVVKSSPPPAPTTKP 993

Qy    733 TAPTTP------------KEPAPTTPKEPAPTTPK--GTAPTTLKEPAPTTPKKRPAPKELAPTT 782
Db    994 LPPPAPVSSPPPVVKSSPPPTTPVSSPPPPPYKKPLPPP---APVS 1050

Qy    783 TKGPTSTTSDKPAPTT--RKETAPTTPK----EPAPTTPKKRPAPTTPETPPTTSEVSTP 836
Db   1051 SPPPVVKSSPPPAPVSLPPPPTTRKPSPPPRTRVSSSPRPVVKCCPPPTLVSSPPPAPKSLPPP 1110

Qy    837 TTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAIKPEMTTTAKDKTTI 896
Db   1111 TPVSSP-------PPEVKSSPPPTPVSSSPPPAPKSSPPPTPVSSPPP------ 1150

Qy    897 ERDLRTTP----ETTTRAAPKMTKETATTTEKTTESKITATTTQVISTTTQDTTPFKITT 951
Db   1151 -ELKSSPPPAPVSSPPSSPAPKSSPPPAPVSLPPPEVKSSPPPAPISSPPPAKSP---- 1203

Qy    952 LKTTTLAPKVTTTFKKTITTTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTK 1011
Db   1204 -----PPPA-----PMSSLPPPVVKSPPPPAPVSSPPPPMKSPPPPAPIS 1242

Qy   1012 KPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSE 1071
Db   1243 SPPPAP--VKPPSLPPPAPVSSPPPAV---------PKKE 1275

Qy   1072 DAGGAEGEGTPHMLLRPHVEMPEVTPDMDYLPRVPNQGIIINPMLSDE 1118
Db   1276 EDSTA-----------------PPAEALPPPSFNDIILPPIMANK 1303
```

```
Db    746 SPT-PPSSSKGAPVPSTGA--------------PPSPKGAPIVPTESSISSSKQVPAEILPSPQ 792
Db    145 SSSSSSSSSTIIWKIKSSKNSAAANRELQKKLKVKDNKKNRIKKKFPTPKPPVVDEAGSGLDN 204

Qy    793 KTPEVTASRLISAVQSPKVDPIMSD------VTPTSPKKISATAVPK---------- 833
Db    886 TLATSIPKVTSPSPQKTPKSVSLKGAPAMTSKKAT---EIAASKDVSPSQ--FPKEVPLL 940

Qy    205 GDFKVTTPDISTTQQHNK-----VSTSPKITTAKPINPRPSLPPNSDTSKETSL-TVNKE 257
Db    834 DIYSATLSLKSVPAVTSLSPPKAPVAPSNEATIVPTELPTSLKNALAAATPKE 885

Qy    258 T------TVEFKETTTNKQTTSTDGKEKTISAKETQSIERKTSAKDLAPTSKVLAKPTPKA 311
Db    998 PEGVTAVPLEIPPCSKKRAPKTAAPKESSATSSSKRAPKTAVSKEIPSKGVTAVPLEISLP 1057

Qy    312 ETTTKGPALTTPKEPTPTTPKEPASTTPKEPATTPKSAPTTPKE-PAPTTKSAPTTPK 370
Db    407 -----TPKEPAPTTPKK-PAPTTPKEPAPTTPKEPTPTTKSAPTTPKE-PAPTTPKE 457

Qy    371 EPAPTTKEPAPTTPKEPAPTTTKEPAPTT---TKSAPT-- 406
Db    648 AAPNTPKEPAPTTPKE-PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTP---- 701

Qy    458 PAPTTAPKPAPTT---PKEPAP---TTPKEPAPTTPKEPSPTTPK----EPAPTTKSA 506
Db   1115 SLAASSQKRSPKTSVPKETPGGVTAMPLEIPSAPOKAPKTAVPKQITPEDAVTLLAGS 1174

Qy    507 PTTTKEPAPTTP-KSAPTTPK-------EPSPTTPKEPAPTTPKEPAPTTPKE-PAPTTPKE 556
Db   1175 PLSPKKASKTAAPKEAPATPSVGVIAVSGEISPSPKKTSKTAAPKENSATLPKRSPKTA 1234

Qy    557 -PKE----PAPTTKREPAP-------TTTKPAPTAPKEPAPTTPKETAPTT 596
Db   1235 APKEIPATSSEGVTAVPSEISPSPPTPASKGVPVTLLIPKGAPNALAE-SPASPKKVPKTA 1293

Qy    702 ----KKPAPKELAPTTTKEPTS-------TTSDKPAPTTPKGTAPTTPKE 740
Db   1294 APEETSTTP-------SPQKIPKVAGPKEASATPSKKIPKTAVPKETSAPSEGVTAVPLE 1347

Qy    741 RAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKKRPAPKELAPTTPKGPTSTTSDKPAPTTP 799
Db   1465 PKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-APKTAAPPSER-VTTVPPEKPA-TPQ 1519

Qy    800 KETAPTTPKEPAPTTPKKRPAPTTPETP-----PPTTSEVSTPTTTK------EPTTIHKS 848
Db   1520 KASGTTTASKVPVPAETQEVAVSSRETPVTPAVPPVKNPSSHKKTISKTIELKEAPATLPS 1579

Qy    849 PDESTPELSAEPTPKALENSPKE-PGVPTTKTPAAIKPEMTTTAKDKTTERDLRTTPETT 907
Db   1580 PTKSPRKIPSSKKAPRI--SAPKEFPASPSIK-PVTT--SLAQTAPPSLQKAPSTTIPKEN 1634

Qy    908 TAAPKMTKETATTTEKTTESKITATTTQVISTTTQDTTPFKITTLKTTTLAPKVTTTKT 967
Db   1635 LAAPAV----LPVSSKSSPAAPARASASLSPATAAPQTAPKEATTIPSCKKAAATETPIET 1690

Qy    968 ITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPPISTKKPKTMPRVRKPKTTPT 1027
Db   1691 STAPSLEGAPKETSE------TSVSKVLMSSP------PKKASSSKRASTLP------- 1730

Qy   1028 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETP 1081
```

```
Qy     85 NPTSPPSSKKAPPSGASQTIKSTTKRSPKPNKKKIKTKKVIESEEITEEHSVSENQDSSS 144
```

```
Db   1731 ----ATTLPSLKEASVLS--------PTATSSGKDSHISPVS-DACSTGTTTP 1770

RESULT 12

ID    Q9LIE8
AC    Q9LIE8;
DT    01-OCT-2000 (TrEMBLrel. 15, Created)
DT    01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS    Arabidopsis thaliana (Mouse-ear cress).
OC    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC    eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX    NCBI_TaxID=3702;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=COLUMBIA;
RA    Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL    Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=COLUMBIA;
RX    PubMed=10907853;
RA    Nakamura Y.;
RT    "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT    Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT    TAC and BAC clones.";
RL    DNA Res. 7:217-221(2000).
DR    EMBL; AP001306; BAB03062.1; -.
DR    InterPro; IPR002965; P_rich_extensn.
PRINTS; PR01217; PRICHEXTENSN.
SQ    SEQUENCE  1480 AA;  147153 MW;  D1AC0C79F155E732 CRC64;


Query Match      11.0%;
Best Local Similarity    28.8%;  Score 811.5;  DB 10;  Length 1480;
Matches  330;  Conservative  60;  Mismatches  539;  Indels  215;  Gaps  51;


Qy     86 PTSPPSKKAPPPSGASQTIKSTTKR----SPKPPN------KKKTKKVIESEEITEEHS 135
          |||                    |:||              ::|    :|| |    ||||
Db     46 PPKPPAVKPHPHPK---PPTIKPPPKRHPHPKPPTVKPHPHPKPPTKPHPHPKPPTKPH- 102

Qy    136 VSENQESSSSSSSSSSSTIWIKSSKNSAANRELQKKLKVK--DNKKNRTKKKPTPKPP 193
          | |             |   | ||  |           |  :  |   || ||||||
Db    103 -------------PHPKPPTIKPPPHPK---PRPHPKPNVKPHPHPKPPTKPHPHPKPP 146

Qy    194 VVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLT 253
          :|                |   ||      :|: ||       ||  | | ||
Db    147 ----------------TKHH---PHPKPPTIKPPPKPPSVKPPPSTPK----- 175

Qy    254 VNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV--LAKPTPKA 311
          |||                    ||:   : ||||     :: ||  : | :|||
Db    176 ------------PPTTNPPSTPQPPTHKPPPCTPTPPVASPMATPTQMPPIATP-PIA 223

Qy    312 ETTTTKGPALTTPKEPTP---TTP--KEPASTTPKEPPTTIKSAPTTPKEPAPTTT---- 362
          ::                           :|    ||    ||||||||||
Db    224 KSPVATPPIATPPITIPPVATPPITTPPIANPPIMPPIATPPVAAPPITNPPIS 283

Qy    363 KSAPTTPKEPAPTTT-------------------KEPAPTTPKEPAPTTTT------ 401
          ||||||||||                        ||||||||||||     -PTTT
Db    284 KPPVTTP----PTTTPPIAKPPIATPPISTPPAAATPPITTLPPAAKPPVAISPIVT 339

Qy    402 KSAPTTP-KEPAPTTPKKPAP---TTPKEPAPTTPKEPTPTTPKEPAPTTTP--KE 457
          ||||||| :|||||||   |||   |||||||||||:|:|||||||||||||  :|
Db    340 --PPVTPIAQPPVATPPIATPPVATPPIATPPISKSPSTPPISESPVATPPIAT 397

Qy    458 PAPTAPKKPAPTTPKEPAPTTPKEPAPTTT---KEPSPTTPKEPAPTTPKEPAPTTTKSAPTTTKEP 513
          ||      |||||||||||||||||||          :|||||||||||||||||:|||||||||
Db    398 PPPAAKPPVATPPIAKSPIATPPVATPPIATPPITTPPIATP---RPVAKPP 454

Qy    514 APTTTKSAPTTPKEP--SPTTTKEPAPTTPKEPAP---TTPKKPAPTTPKEPAPTTPKEP 568
          |||||:|||||||||  :|||||||||||||||||   ||:|| |||||||||||||:
```

```
Db    455 VETPPIATPPTAKPPILSTPPISKPPVATPPAATPPITTPPPEVKPPVATPPLAIPPVAKPP 514

Qy    569 --APTTTKKPAPTAP--KEPAPTTPKETAPTT---------PKKLTP--TTPEKLAPTTPEKP 616
          :| |||||| : |||  |||||||||:||||           |||||:  |||:|||||||
Db    515 VVTPPIATPPIATPPIAKSPVATPPITTPPTATPPVVTPPIATPPVTPTTTPPTATPPVAKPP 574

Qy    617 APTTPEELAPTT-PEEPTTTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKEP 672
          |  || ::|| ||  ||  || :|||||||         || ||||||||||||||||
Db    575 VATPPIATPPIAKPPISTPPISKSPVATPPAATPPITTPPPAKPPVATPPIATPPIAKP 634

Qy    673 APTTTPKETAP----TTPKGTAPTTT---LKEPAPTTPKKPAPKELAPTTTKEPTSTTTSDKPAP 727
          |||||||||       |||| ||||      ||||||||  ||||||||||||:||||
Db    635 VATPPTATPPIATSPVAKPPVATPPIKTPPPAKPPVAIPPIATPPVAKPPVAT---PPTA 691

Qy    728 TTPKGTAP-TTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGP 786
          |||| ||  |||||||||||:|||||||||||||:|||||||||  |||||||||||||
Db    692 TPPIATPPIATPPVTTPPIATPPVATPPIAKPPITI-PPTATPPVAMPPIATPPTATKPP 749

Qy    787 TST--TSDKPAPTTPKETAPT--TPKEPAPTTPKKPAPTTP-----EPPPPPTTSEVSTPT 837
          | :| |||||||||| |||||  ||||||||||| |||||      :|||||||||
Db    750 IATPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPIATPPVATPPIAKPPVATPP 809

Qy    838 TTTKEPTTIH--KSPDESTPELSAEPTPKRALENSPKEPGVPTTKTPAATKPEMT-TTAKDK 894
          ||||||||  ::||| : |||||||||||:|||||||| |||||||||||::
Db    810 TTAPPTATPPVAKPPVATPPIATPPTAKPPILTPPISKPPVATPPATPPIAKPPVATPP 869

Qy    895 TTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTTPKITTLKT 954
          || ||| || : :|||| | |   ||      | ||  | ||||||||||||
Db    870 VATPPIATPP--IAKPVATPPVATPPIATSPVAKPPVAIPPIKTPPPAKPPVAIPPIAT 927

Qy    955 TTLAPKVTTTKKT----ITTTEIMNKPEETAKPDKDRATNSKAATTKPPQKPTKAPKKPPTST 1010
          |:|||||||||:    |   |||| ||:|  |||  ||   ||||||||||||||||:|
Db    928 PPVAKPPVATPPTATPPIATSPIATPPVVT----PPIATSPVATPPIAKPPTTTP--PTAT 982

Qy   1011 KKPKTMPRVRKPKTTPTPPRKMTSTMPELNP------------TSRIAE------AMLQTTR 1054
          ||| ||    |||| |  ||  :: |:||          :| ::||      :|||||
Db    983 -PPVAMPPIATP---PTAKPPVATPPIANPPVEKPPVATPPIAKPPVLPPIAKPPVETS 1038

Qy   1055 PNQTPNSKL--VEVNPKSEDAGGAEGETPHMLLRPHVFEMPEVTPDMDYLPRVPNQGIIIN 1112
          || || ::  |  |:   |     ||: |    || |:|||
Db   1039 PIATPPPIATPPVAIPP------------VVKPPVAIPPITKPPVATPPVTNPPIAMP 1083

Qy   1113 PMLS 1116
          |:::
Db   1084 PIVT 1087


RESULT 13

ID    Q41805
AC    Q41805;
DT    01-NOV-1996 (TrEMBLrel. 01, Created)
DT    01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    EXTENSIN-LIKE PROTEIN PRECURSOR.
OS    Zea mays (Maize).
OC    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC    Panicoideae; Andropogoneae; Zea.
OX    NCBI_TaxID=4577;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=B73; TISSUE=POLLEN;
RA    Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL    Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; Z34465; CAA84230.1; -.
DR    Mendel; 14346; Zeama;2368;14346.
DR    InterPro; IPR001611; LRR.
DR    InterPro; IPR003592; LRR_out.
DR    Pfam; PF00560; LRR; 3.
DR    SMART; SM00370; LRR; 4.
KW    Signal.
```

PRELIMINARY;   PRT;   1480 AA.

PRELIMINARY;   PRT;   1188 AA.

```
FT   SIGNAL         1     27
SQ   SEQUENCE   1188 AA;  120980 MW;  2C77C7F8D7130149 CRC64;
     POTENTIAL..
```

Query Match
Best Local Similarity   29.9%;   Score 784; DB 10; Length 1188;
                                 Pred. No. 1e-42;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

```
Qy    188  PTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPIPNRPSL----PPN  243
           ||||  ||                                   ||:|| |||||    |||
Db    460  PTPHSPPAD--------DYVPPTPVPVGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQ   510

Qy    244  SDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKITSAKETQSIEKTTSAKDLAPTSKV 303
           :                                                    |||
Db    511  APVG---------------------------------SPPPPVKTTSPPAPIG---SPSPPPVSVV 541

Qy    304  -----LAKPTPKAETTIKGPALTTPKEPTPTTPKEPASTTKEPTPTTIKSAPTTPKEPA   358
                :| | : | : |    | |||  | ::||||| |:||  :|||
Db    542  SPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVASPPPPVKSP--PPPTLVASPPPPVKSPP  599

Qy    359  PTTTKSAPTTP--KERAPTTTKEPAPTTTKEPAPTTTKEPAPTTIKSAPTTP--KERAPTT 415
            :: ||: || :  |||     | :||||: | ::||||  | |||| |||    | |||
Db    600  PPAPVASPPPPVKSPPPPTPVASSPPPPAVASSPPPPTPVSSPPPPPEKSPPPP        659

Qy    416  PKKPAPTTPKEPAPTT------PKE---PTPTTPKEPAPTTTKEPAPTTTKEPAPTTTKKP 466
           | | || | :|||||      |:|   |:| | :| :|||| |||||| |:|||||| |
Db    660  PAKSTPPPEEYPTPPTTSVKSSPPPEKSLPPPTLIPSPPPQEKPTPPSPTPSKP-PSSPEKP 718

Qy    467  APTTPKEPAPTTPKEPAPTTTKEPSPTTTKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPK  526
           :| | :|| | : :|| : | ||:| |: ||| | : ||| |  :|| |  :|| |
Db    719  SP--PKEPVSSPPQTP----KSSPPAPVSSPPTPVSSSPPALAPVSSPPPSPVKSSP---- 768

Qy    527  EPSPTTTKEPAPTTPKKPAPTTPKEPAPTTTKKPAPTTKEPAPTTTTKPAPTAPKEPAP    586
            | :| | |:|| | | || | |:|| |  :|| |   ||| | : | ||  |:|||
Db    769  PPAPLSSPPPAPQVKSSPPPVQVSSP-PPAPKSSPPLAPVSSPPQVEKTSPPPAPLSSPP  827

Qy    587  TTPKETAP---TTP---KKLPTTPEKLAPTTPEKPAP------TTTEELAPTTPEEPT     633
            | | ::|   | |   | :| :|:| :| | :| ||      |  ::|| | :|
Db    828  LAPKSSPPSPHVVSSPPPATPKSSPPAHVSSPPPPVSSPPPA---PKSSPPPAP       885

Qy    634  PTTPEEPAPTTPKAAAPNTPKE-PAPTTPKEPAPTTPKETAPTTTKPAPTTTKGTAPTT    692
           |:: |:| | : |  | | |: ||| | |:|| | |:|:|| |  || | : |:|||
Db    886  PTTVISP-PSEPKSSPPPIPVSLPPPIVKSSPPPAMVSSP-PMTPKSSPPPVVVSSPPPI  943

Qy    693  LKE---PA---PTTPKK---PAPKELAPTTTKE---PTSTTSDKPAPTTPKGTAPTTP    738
            |:   || |: :|| | ||  :| :||  :|   |:  ||| | |:| |:|| | |
Db    944  VKSSPPPAPVSSPPATPKSSPPAPVNLPPPEVKSSPPPPVSSPPPA---PKSSPPPAP  1000

Qy    739  -KERAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPT  797
            ::| | | :|| | | :|| | |:|| | |:|| :| :|||  :| |:|:  ||| | |
Db   1001  MSSPPPPEVKSPPPPAPVSSPPPPVKSPPPPAPVSSSPP----PPVKSPPPPAPVSSPPPPV 1057

Qy    798  -TPKETAPTTTPKEPAPTTPKPAPTTPETPPTTSEVSTPTTTKEPTTIHKSPDESTPEL   856
            :| :||:  :| :|| | | | |: | ||  | ||:| | |:|  | :|| :|:| |
Db   1058  KSPPPPAPISSPPPPVKSPPPPAPVS--SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS  1115

Qy    857  SAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPET           906
           |:|||: | :|| :| |: | | |  ::| :|  ||  || |:| |:|:
Db   1116  SPPPAP-----VKPPSLP-PPAPVSSPPPVVTPAPPKKEEQSLPPPAES           1158
```

```
RESULT   14
Q9VEL9
ID   Q9VEL9        PRELIMINARY;   PRT;   2112 AA.
AC   Q9VEL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG4090 PROTEIN.
GN   CG4090.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003716; AAF55402.1; -.
DR   FlyBase; FBgn00038492; CG4090.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF01607; Chitin_bind_2; 10.
DR   SMART; SM00494; ChtBD2; 11.
DR   PROSITE; PS00022; EGF_1; 1.
SQ   SEQUENCE   2112 AA;  219547 MW;  B9101B5551A5D36 CRC64;
```

Query Match
Best Local Similarity   10.6%;   Score 782.5; DB 5; Length 2112;
                        27.1%;   Pred. No. 2.4e-42;
Matches 310; Conservative 106; Mismatches 369; Indels 361; Gaps 49;

```
Qy     54  FERGRECDCD------AQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAPPSGASQTIK   106
           |:  | | |       | || | || | :|| |  :||:  ||::| |||:|:| | |
Db    969  FKPAEKCESEETFLADNENCSKFYRCVDNGKGGFFKV-SFTCPPNTLWDPEANSCNHPDQ 1027

Qy    107  STTKRSPKPPNKKKTKKVI-----ESEEITEEHSVSENQESSSSSSSSSSSSSTIWKIKSSK 162
           ::|| |||  | |||: :|     |:||  |  |  |:|::|||||||||||||
Db   1028  IQTK-------PLKCKKVVSQGGSSSNSTSNSSSSGNSGGSSNSGSSSNSGSSSSS----SSS 1074

Qy    163  NSAANRELQKLKVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVITPDTSTTQHNKV  222
           ||                                    |:||::|
Db   1075  NS--------------------------------GSSSNTGSSSNSGASSSGSSSNOG    1101

Qy    223  STSPKITTAKPINPRPSLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTS 282
           |:|                                                     |:|
Db   1101  SSS---------------------------NSGSSSGSNSSGNQSTS------SSTSSSSSSSSNNNQGSS 1138
```

```
RESULT 15
Q9XDH2

ID   Q9XDH2      PRELIMINARY;    PRT;    763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH MUCIN HOMOLOG.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang Y., Moreno C., Singh M.;
RT   "Cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   InterPro; IPR002965; P_rich.
DR   InterPro; IPR003882; Pistil_extensin.
DR   PRINTS; PR01222; ATROPHIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE   763 AA;   75034 MW;   39168EC45A5916F8 CRC64;

Query Match    9.6%;   Score 715;   DB 2;   Length 763;
Best Local Similarity   28.2%;
Matches  248;  Conservative  39;  Mismatches 361;  Indels 232;  Gaps 39;   Pred. No. 1.9e-38;
```

Left column:

```
Qy   283  AKETQSIEKTSAK-DLAPTSKVLAKPTPKAETTTK------------------------GP----  318
Db  1139  SSSSSSSSSTSSKPNPSETCKVNGQFIGDRSDCAKFYRCVDNDRGGEFNMYFSCGPGTVW       1198

Qy   319  ------------------ALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTT      361
Db  1199  DAQMQACNHAWAVKECGGIAPPTTSPTTTSR-PTTASTSRPSDQTSTSRPTGP----PTT       1253

Qy   362  TKSAPTTPKEPAPTTPKEPAPTTPKEPAPTT------------------------------      392
Db  1254  ARPVTARPTTSSPTTASSSQTTSPVTQAPNTDGKCRSEGFMADPNNCSKFYRCVRNNKGG       1313

Qy   393  -TKEP-------APT---------------------------------------------       431
Db  1314  FTSIPFQCGAGTVWDQDLQTCNHNFNNCSTGTESTTPKPPC--EPATNGTTATSTSSTTT       1371

Qy   432  PKEPPTTPKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTKPEPS              491
Db  1372  P--PPTTDLPPTSTTGLP-PTTTTELPPT                                      1421

Qy   492  PTTPKEPAPTTIKSAPTTPKEPAPTTPKEPAPTTPK---EPSPTTTKEPAP               538
Db  1422  TTTGLPPTTTTGAQPTTTTLSSETETSTVTTSPESTTQPPSTTTMKLPAGTECTGEGY        1481

Qy   539  ------------TTPK---------------EPAP----TTPKKPAPTT                 556
Db  1482  MADPEDCRKYRCINAGASYRKNFTCPKGTGWNEEVQTCDYVENIPRCSKLPAEPITTT        1541

Qy   557  P---KEPAPTTPKE-PAPTTAPKPEAPTTPKETAPTTPKETATPTTPKKLTPTTPEKLAPT      611
Db  1542  PSEESKDPGSTTPQSTDEPTTVIKPI-TKPTE-EPSTEKPQKPTTQYPEKPTTTEE--PE      1597

Qy   612  TPEKPAPTTPEELAPTTPEEPTPTTPKAAAPNTPKEPAPTTPKEPAPTT-PK              670
Db  1598  KPQKPT-TTEYPQKPTTTEEPTPEKPQKPTT-TTEYPQKPTTTEYPQ                   1655

Qy   671  EPAPTTPKETA------PTTPK---GTAPTTLKEPAPTTPKETAPTTPKEPAPTTPKKPAPKELAPTTPKEPISTTSD   723
Db  1656  KPTTEEPTTTSIPGYNPTTTTSVPGYNPTTTTPIPVETTTSTPGYK---                 1711

Qy   724  KPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTPKEPAPTTPKK---PAPKE        777
Db  1712  LPSTTTDAIQEPPTSKKPEPPTTTESPESSTPEGSVTTLQPEPQPNYNCSSEGFFPDPED      1771

Qy   778  LA------PTTTKGPTSTTSDKPAPTTPKETAPTTTKPEPAPTTPKKPAPTTPETPPPTTS      831
Db  1772  CSRYYRCVDAAANGKYQVYAFKCGKGTVWDISTETCNYADQVSGN                     1820

Qy   832  EVSTPTTTKEPTTIHKSPDESTPELSAAEPTTPKALENSPKEPCGVPTTKTPAATKPEMTTTA    891
Db  1821  QTTTPGTTTEPGTTESTTSSGKP---ETTSKAPEN                               1863

Qy   892  KDKTTERDLRTTPETTTPAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTTPFKITT    951
Db  1864  ----SSPETTTT---VASETTTTTSGT                                       1898

Qy   952  LKTTTLAPKVTTTKKITTTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTK     1011
Db  1899  -ETTTIAGEESTSKSPTTTE                                              1929

Qy  1012  KPKIMP                                                            1017
Db  1930  CPETGP                                                            1935
```

Right column:

```
Qy   307  PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAP      366
Db     3  PVP--------APRLAPLPPAPPAPAEPKSKPFPPAP-------PAPPCWMLVSAAP          46

Qy   367  TTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTTKKPAPTTPKE              426
Db    47  PCP--PAP----PAPPRKPKSKAPPPVPPAPPAPARELAPPLP--PEAPRE               90

Qy   427  PAPTTPKEPTPTT----PKEPAPTTPKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK    480
Db    91  SRPALPPCPPPVVPIPDPEPERAAPVPPAPNSPPPFPPPPPAPKFVPAPPVP-PVNNSPP     148

Qy   481  EP--APTTTKEPSPTTTPKEPAPTTTKSAPTTPKEPSPTTTKEPAP                   538
Db   149  FPPPPPAALNPPAP-----PAPPLANSPPLPPAPPTPAGT--PPAAPWPPPVPAAAPKSKPA   201

Qy   539  TTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAP----------TTTKKPAPTAPKEPAPTTPKETAP   594
Db   202  SPPRPPAD-----PMPAPAMERPPLPPVPPDPDISKEPPAPPAPPIPPAPPVPIPPVPVPPLP   256

Qy   595  TTPEKPAPTTPEELAPTTPEEPTPTTPEKLAP-------                           644
Db   257  PVPNKIPPAPP---AEPVAVAAVLVAPCPPLPPLPNNHPAPPAAPVPGVPLAPLPNSHP       313

Qy   645  ---PKAAAAPNTPKEPAP                                                658
Db   374  TTTTTALPAPIPPLPPLPPLPINTAVPPIPLPPPVTALAPPLPLPISGGVPPAP--          431

Qy   709  LAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPA-------                         762
Db   432  --PIPPGKPWTTPLAPAPPEEK-TVPVLPPGPSCPPSEKPNPPAPPEPEPERKSSPALPP     488

Qy   763  KEPAPTTP------KKPAPKELAPTTTKGP-TSTTSDKPAPTTPKET---                815
Db   489  APPAPSMPSAVRVPPSPPIPPAPPAAPRASMPALPPAPSPPPATRLCPPLPPSPPAPNSP     548

Qy   816  KKPAPTTPTETPPPTTTSEVSTPTTTKEPTTIIKSP----DESTPELSAEPPPKA--LENSP   869
Db   549  --PAPPAPPTPPKLLS--ANPCCPPVPPAPNRPAPPAPPAPPELPAPPDPPTPPVANSP      604

Qy   870  KEPGVPTTKIPAAATKPEMTTTAKDKTTERDLRTTPETTTPAAPKMTKETATTTEKTTESKI   929
```

```
Db   605 --PAPPAPPAPPSALPFVNPPA----------------------PPTPAAPK-----------SRPAL 637
Qy   930 TATTTQVTSTTTQDTTPFKIITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATN 989
         |      : | |||                                  | |
Db   638 PAAPPAPPAPPVRATTP-------------------------------PPAPPAPPAPN 665
Qy   990 SKATTPKPQKP-----TKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTM-----PEINP 1040
         |  |  |  | |     | | :  | | : :                        | :
Db   666 SMALPPAPPDPPIPLLATPPAPPAPPLPMSPPAPPLPPAAPDPPAPPLTINQPPSPPLAP 725
Qy  1041 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET 1080
         | | |      ||        : | : :|:|
Db   726 VPGAPLAPLPINGRPVFARKNSLI--------GSSSGDT 756
```

Search completed: April 26, 2002, 16:22:02
Job time: 507 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:22:18 ; Search time 153.5 Seconds
           (without alignments)
           1013.901 Million cell updates/sec

Title: US-09-556-246-1_copy_200_1263
Perfect score: 5698
Sequence: 1 VKDNKKNRTKKKPTPKPPVV..........VAALSTAKYKWNPESVYFFK 1064

Scoring table: BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_17:*
      1:    sp_archea:*
      2:    sp_bacteria:*
      3:    sp_fungi:*
      4:    sp_human:*
      5:    sp_invertebrate:*
      6:    sp_mammal:*
      7:    sp_mhc:*
      8:    sp_organelle:*
      9:    sp_phage:*
     10:    sp_plant:*
     11:    sp_rodent:*
     12:    sp_virus:*
     13:    sp_vertebrate:*
     14:    sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5698	100.0	1404	4	Q92954	Q92954 homo sapien
2	5689	99.8	1404	4	Q9bx49	Q9bx49 homo sapien
3	2519	44.2	1054	11	Q9jm99	Q9jm99 mus musculu
4	1048.5	18.4	401	6	Q77765	Q77765 bos taurus
5	967.5	17.0	1079	5	Q94A57	Q94A57 caenorhabdi
6	936.5	16.4	1049	5	Q9n4s7	Q9n4s7 caenorhabdi
7	865.5	15.2	1795	5	Q917s1	Q917s1 drosophila
8	853	15.0	1489	10	Q76894	Q76894 drosophila
9	808.5	14.2	1315	10	Q96449	Q96449 phytophthor
10	802	14.1	1274	5	Q9spm0	Q9spm0 zea mays (m
11	800.5	14.0	1180	10	Q20007	Q20007 caenorhabdi
12	784	13.8	1188	10	Q9lie8	Q9lie8 arabidopsis
13	776	13.6	2187	11	Q41805	Q41805 zea mays (m
14	739	13.0	2112	5	P70670	P70670 mus musculu
15	715	12.5	763	2	Q9vel9	Q9vel9 drosophila
16	707.5	12.4	555	10	Q9xdh2	Q9xdh2 mycobacteri
17	671	11.8	1151	13	Q9fpq6	Q9fpq6 chlamydomon
18	667.5	11.7	4880	11	Q9jlt1	Q9jlt1 rattus norv
19	667.5	11.7	5085	11	Q9jks6	Q9jks6 rattus norv

RESULT 1
Q92954

ID   Q92954   PRELIMINARY;   PRT;   1404 AA.
AC   Q92954;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEGAKARYOCYTE STIMULATING FACTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Blood 78:279-279(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA   FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA   Jacobs K., Turner K.;
RL   (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL   Mosher D.F. (eds.);
RL   Biology of vitronectins and their receptors., pp.45-52,
RL   Elsevier Science Publishers B.V. (1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
EMBL; U70136; AAB09089.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   InterPro; IPR002400; GF_Cysknot.
DR   Pfam; PF00045; hemopexin; 2.
DR   Pfam; PF00086; Somatomedin_B; 2.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR00438; GFCYSKNOT.

ALIGNMENTS

20	663.5	11.6	2284	5	Q9VPG1	Q9vpg1 drosophila
21	660.5	11.6	1229	5	Q9N435	Q9n435 caenorhabdi
22	659.5	11.6	1229	5	Q94185	Q94185 caenorhabdi
23	653.5	11.5	1514	5	Q9GUW7	Q9guw7 leishmania
24	647	11.4	4833	11	Q9QYX6	Q9qyx6 mus musculu
25	647	11.4	5038	11	Q9QYX7	Q9qyx7 mus musculu
26	643	11.3	2089	4	Q14676	Q14676 homo sapien
27	633	11.1	489	10	Q41707	Q41707 vigna ungui
28	632.5	11.1	7962	4	Q10465	Q10465 homo sapien
29	632	11.1	761	10	Q9ZQI0	Q9zqi0 arabidopsis
30	632	11.1	6632	5	Q17362	Q17362 caenorhabdi
31	629	11.0	3507	5	Q23587	Q23587 caenorhabdi
32	628.5	11.0	1612	5	Q9VQ02	Q9vq02 drosophila
33	626.5	11.0	990	13	Q91803	Q91803 xenopus lae
34	625	11.0	6642	5	O01761	O01761 caenorhabdi
35	624.5	11.0	2768	5	Q9VC00	Q9vc00 drosophila
36	622.5	11.0	839	2	Q9RX57	Q9rx57 deinococcus
37	622	10.9	3570	4	Q99552	Q99552 homo sapien
38	620.5	10.9	971	5	Q9XVS4	Q9xvs4 caenorhabdi
39	615	10.8	2344	5	Q9N3Y8	Q9n3y8 caenorhabdi
40	607.5	10.7	801	5	O23635	O23635 caenorhabdi
41	607.5	10.7	924	12	Q99307	Q99307 epstein-bar
42	605	10.6	3893	5	Q27929	Q27929 drosophila
43	600.5	10.5	1893	5	Q9NKC9	Q9nkc9 drosophila
44	598.5	10.5	409	10	Q9SBM1	Q9sbm1 volvox cart
45	577	10.1	956	10	Q9LJ64	Q9lj64 arabidopsis

```
DR   PRINTS; PR000022; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS005024; SOMATOMEDIN_B; 2.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
SQ   SEQUENCE   1404 AA;  151090 MW;  AABD7AD19B35F4F6 CRC64;

Query Match             100.0%;  Score 5698;  DB 4;  Length 1404;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1064;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTDTSTTQHNKVSTSPKITTAKPIN 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  200 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTDTSTTQHNKVSTSPKITTAKPIN 259

Qy   61 PRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  260 PRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 319

Qy  121 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPTPTTIKSAPTTPK 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  320 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPTPTTIKSAPTTPK 379

Qy  181 EPAPTTKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  380 EPAPTTKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTT 439

Qy  241 PKKPAPTTPKEPAPTTPKEPTPTTTKEPAPTTTKEPAPTTPKEPA 300
        ||||||||||||||||||||||||||||||||||||||||||
Db  440 PKKPAPTTPKEPAPTTPKEPTPTTTKEPAPTTTKEPAPTTPKEPA 499

Qy  301 PTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPSPTTTKE 360
        |||||||||||||||||||||||||||||||||||||||||||
Db  500 PTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPSPTTTKE 559

Qy  361 PAPTTPKEPAPTTPKKPAPTTPKEPAPTTTKPAPTAPKEPAPTTPKETAPT 420
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  560 PAPTTPKEPAPTTPKKPAPTTPKEPAPTTTKPAPTAPKEPAPTTPKETAPT 619

Qy  421 TPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  620 TPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 679

Qy  481 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  680 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 739

Qy  541 EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTTKEPAPTTTKEPAPTTPKEPAPTT 600
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  740 EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTTKEPAPTTTKEPAPTTPKKPAP 799

Qy  601 KELAPTTPKGPTSTISDKPAPTTPKERAPTTPEPPPPTTSEVST 660
        ||||||||||||||||||||||||||||||||||||||||
Db  800 KELAPTTKGPTSTISDKPAPTTPKETAPTTPEPPPPTTSEVST 859

Qy  661 PTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKT 720
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  860 PTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKT 919

Qy  721 TERDLRTTPETTTAAPKMTKETATTTEKTTEESKITATTTQVTSTTTQDTTPFKITTLKTT 780
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  920 TERDLRTTPETTTAAPKMTKETATTTEKTTEESKITATTTQVTSTTTQDTTPFKITTLKTT 979

Qy  781 TLAPKVTTTKKITTTIEIMNKPEETAKPKDRATNSKRATTTPKPQKPTKAPKKPTSTKKPKT 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  980 TLAPKVTTTKKITTTIEIMNKPEETAKPKDRATNSKRATTTPKPQKPTKAPKKPTSTKKPKT 1039

Qy  841 MPRVRKPKTPTPPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG 900
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1040 MPRVRKPKTPTPPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG 1099

Qy  901 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDEFNICNGKPVDGLITLRN 960
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1100 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDEFNICNGKPVDGLITLRN 1159

Qy  961 GTLVAFRGHYFWMLSPFSPSPSARRITEWNGIPSPIDTVPTRCNGEGKTFFFKDSQYWRF 1020
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1160 GTLVAFRGHYFWMLSPFSPSPSARRITEWNGIPSPIDTVPTRCNGEGKTFFFKDSQYWRF 1219

Qy 1021 TNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYFFK 1064
        ||||||||||||||||||||||||||||||||||||||||||
Db 1220 TNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYFFK 1263
```

```
RESULT 2
Q9BX49

ID   Q9BX49       PRELIMINARY;      PRT;   1404 AA.
AC   Q9BX49;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).
GN   BG174L6.2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133553; CAC36090.1; -
SQ   SEQUENCE  1404 AA;  151076 MW;  782A11746B3FDEE5 CRC64;

Query Match             99.8%;  Score 5689;  DB 4;  Length 1404;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches 1062;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    1 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTDTSTTQHNKVSTSPKITTAKPIN 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  200 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTDTSTTQHNKVSTSPKITTAKPIN 259

Qy   61 PRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  260 PRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 319

Qy  121 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPTPTTIKSAPTTPK 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  320 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPTPTTIKSAPTTPK 379

Qy  181 EPAPTTKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  380 EPAPTTKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTT 439

Qy  241 PKKPAPTTPKEPAPTTPKEPTPTTTKEPAPTTTKEPAPTTPKEPA 300
        ||||||||||||||||||||||||||||||||||||||||||
Db  440 PKKPAPTTPKEPAPTTPKEPTPTTTKEPAPTTTKEPAPTTPKEPA 499

Qy  301 PTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPSPTTTKE 360
        |||||||||||||||||||||||||||||||||||||||||||
Db  500 PTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPSPTTTKE 559

Qy  361 PAPTTPKEPAPTTPKKPAPTTPKEPAPTTTKPAPTAPKEPAPTTPKETAPT 420
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  560 PAPTTPKEPAPTTPKKPAPTTPKEPAPTTTKPAPTAPKEPAPTTPKETAPT 619

Qy  421 TPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  620 TPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 679

Qy  481 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  680 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 739
```

```
Qy    541 EPTSTSTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 600
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 EPTSTTCDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 799

Qy    601 KELAPTTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKEPAPTTPETPPPPTSEVST 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    800 KELAPTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPPTSEVST 859

Qy    661 PTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTKPAATKPEMTTAKDKT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    860 PTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTKPAATKPEMTTAKDKT 919

Qy    721 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTQDTTPFKITTLKTT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    920 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTQDTTPFKITTLKTT 979

Qy    781 TLAPKVTTTKKTIITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSKKPKT 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    980 TLAPKVTTTKKTIITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSKKPKT 1039

Qy    841 MPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1040 MPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG 1099

Qy    901 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1100 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN 1159

Qy    961 GTLVAFRGHYFWMLSPFSPPSSPPSSPARRITEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRF 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1160 GTLVAFRGHYFWMLSPFSPPSSPPSSPARRITEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRF 1219

Qy   1021 TNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYFFK 1064
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1220 TNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYFFK 1263
```

```
RESULT 3
Q9JM99
ID   Q9JM99          PRELIMINARY;   PRT;   1054 AA.
AC   Q9JM99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE   PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN   PRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ikegawa S., Nakamura Y.;
RT   "a novel mouse gene highly homologous to a human gene encoding
RT   megakaryocyte stimulating factor precursor and cartilage superficial
RT   zone protein.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034730; BAA23210.1; -.
DR   MGD; MGI:1891344; Prg4.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001212; Somatomedin_B.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR01033; Somatomedin_B.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ   SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;
```

```
Query Match    44.2%;   Score 2519;   DB 11;   Length 1054;
Best Local Similarity    49.6%;   Pred. No 2.4e-160;
Matches  533;  Conservative  38;  Mismatches 136;  Indels 368;  Gaps 20;

Qy    1 VKDNKNRFKKKPPPKPPVVDEAGSGLDNGDFKVT----TPDTSTSTQHNKVSTSPKITTAKP 58
        |||||| |||  |||  : ||| ||   | |::|    | ||  |||  |||  || |||||
Db  195 VKDNKKNTPKKKPNPEPPAVDEACSGLDNGEFKLTPPPPDPPPTTPHSKVATSPKTTAAKP 254

Qy   59 INPRSLPPENSDTSKETSLTVNKETIVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTS 118
        :|   || ||:| :|||   |  |  |: |:|  :  ||  |  :: :|| |  | | ||
Db  255 VTPKPSLAPNSETSKEASLASNKETTVETKETTATNKQSSA -SKKKTTSVKETRSAEKTS 313

Qy  119 AKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAPTTPTTIKSAPTT 178
        ||  |||: :|||| |||  :|||||| ||  : || ||||||    ||: ||||
Db  314 DKDV----------------EPTSTTPK ------NSAPTT 331

Qy  179 PKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTLKEPAPTTPKEPAP 238
        ||||||||:  ||||||||||||||||| |||||| |||||||||||||||| 
Db  332 TKKPV-TTTKESKFLP----------------LPQEPEPTTPKEPEPTTPKEPEP 376

Qy  239 TTPKKPAPTTPKEPAPTTPKETAPTTPKEPAPTTPKEPAPTTKKPAPTTPKE 298
        :|| |||||||||| |:|||||||||| ||| ||||||| |||||| |||||
Db  377 TTPKEPEPTTPKEPEPTTPKEPAPTTPKEPEPTTPKEPGPTTPKEPEPTTKE 436

Qy  299 RAPTTPKEPAPTTKEPSPTTPKEPAPTTTKEPAPTTKSAPTTTKEPAPTTPKEPSPTTT 358
        |  ||||||||| ||  :||||||||||  ||| :|| ||||||||:|| ||||  |||
Db  437 PEPTTTKEPESTTPRKEPEPTTPKEP----------------EPTTPKEPEPTTL 474

Qy  359 KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKKPATPTAPKEPAPTTPKETA 418
        ||| ||||||||||| :|: :| :||| ||||||||   |:| :| :|||  |||||||
Db  475 KEPEPTTPKEPEPTTPKEPEPTTPKEP--------- ----------------- 512

Qy  419 PTTPKKLFLPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTP 478
        |||||: ||||| ||  |||||||||||||||||||||||  :|||  |:|||||   |||
Db  513 -----------------------------------------------------TP 514

Qy  479 KEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT 538
        ||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  515 KEPEPTTPKEPEPTTPKEPEPTTPKEP--------------KEPEPTTPKKP- 549

Qy  539 TKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPEPTTP--------------- 581
        |||||||||||||||||||||||||||||||||||||||||
Db  550 TKKPV-TTTKESKFLP----EPTTPKEPVPTTPKEPEPTTP------KEPEPTTPKEP 598

Qy  599 APKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPPTSEV 658
        :|| |||||||||||||||||||||||||||||||||||||||  |:|| ||| :||||
Db  599 EPTTPKEPVPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP ------------ 617

Qy  582 EP----------TTIKEPEPTTPKEPEPTTPKEPEPTTPKKPEPTT--------- 617
        |: ||:|||||:||||||
Db  582 EP------------TTIKEPEPTTPKEPEPTTPKEPEPTTPKKPEPTT------ 617

Qy  659 STPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTKPAATKPEMTTAKD 718
        |:  |||:|| || ||  | :|| :| |||: || || |  |||||:|| : ||
Db  659 EVPDHEDADGGEGEKP-LIPGPVLFPAIPGTDLLAGRLNRGININPMLSDETNLCNG 718

Qy  719 KTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTQDTTPFKITTLK 778
        :|:|:||||||||||| |:|| :||||| ||  |:  | | :  | | || || ||||
Db  719 -----TSPKT------ ------------------TTLK 626

Qy  779 TTTLAPKVTTTKKTIITTTEIMNKPEETAKPKDRATNSKAT- 829
        ||||||||||||:|::|||| |
Db  627 ATTLAPKVTRAPAE---EIQNKPEETTPASEDSDDSKITLKP 681

Qy  830 KKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLV 889
        |||  |||||||| |||||||||||||||||||||||||||||||||||||||:||||||
Db  682 KKPTSTKKKPKT-PKTRKPKTTPAPLKTTISAPTPELNTTP--LEVMLPTTTIPKQTPNPETA 738

Qy  890 EVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNG 949
        |||||| |:| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  739 EVNPDHEDADGGEGEKP-LIPGPVLFPAIPGTDLLAGRLNRGININPMLSDETNLCNG 797

Qy  950 KPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSSPPSSPARRITEVWGIPSPIDTVFTRCNCEGKT 1009
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db  798 KPVDGLTTLLVAFRGHYFWMLNPFRPPSSPPRRIIEVWGIPSPIDTVFTRCNCEGKT 857

Qy 1010 FFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVYFFK 1064
```

```
Db    858 FFFKDSQYWRTNDVDPGYPKQIVKGFGGLTGKIVAALSIAKYKDRPESVYFFK 912
```

RESULT 4

```
ID   O77765       PRELIMINARY;     PRT;   401 AA.
AC   O77765;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ARTICULAR CARTILAGE;
RA   Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA   Aydelotte M.B.;
RT   "Immunodetection and Partial cDNA Sequence of the Proteoglycan,
RT   Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT   Joints.";
RL   J. Orthop. Res. 0:0-0(1998).
DR   EMBL; AF056218; AAD13404.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   Pfam; PF00045; hemopexin; 2.
DR   PROSITE; PS00024; HEMOPEXIN; 2.
DR   SMART; SM00120; HX; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   401 AA;  44952 MW;  86147CC9AFBB73D7 CRC64;
```

```
Query Match             18.4%;  Score 1048.5;  DB 6;  Length 401;
Best Local Similarity   76.7%;  Pred. No. 1.3e-62;
Matches 198;  Conservative  20;  Mismatches  37;  Indels  3;  Gaps  3;
```

```
Qy   808 PKDRATNSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRK-MTSTMPELNPT 866
         |||||:||||:||||||||||||||||||||:|||||:||||||||||| |||:|||| ||
Db     5 PKGRATNSQVTTPKPQKPTKAPKKPTSTKKPRT-PRVKRPKTTPTPKTTTSAMPEPTPT 63

Qy   867 SRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYL 926
         : |||||||||  | ||:|||  |||:|| :||  |||  |||||:|||||||||||  :
Db    64 S-LPEAMLQTTTRPTPTNSEIIDVNSENEDGDAAEGEKPHMIFRPPVLTPIVIPGTEII 122

Qy   927 PRVPNQGIIINPMLSDETNICNGKPVDGLITLVAFRGHYFWMLSPFSPPSSPARRI 986
         |:||| |||||||||||:|||||||||||||||||||||||||:|||:||||||||
Db   123 VRGPSQGFGINPMFSDEINLCNGRPVDGLITLLRNGTLVAFRGHYFWMLTPFTPPPPRRI 182

Qy   987 TEVWGIPSPIDTVTTRCNCEGKGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVA 1046
         ||||||||||||:||||||||||||||||:||||||||:||||:||||:|||||||:|||
Db   183 TEVWGIPSPIDTVTTRCNCEGKTFFFKGSQYWRFTNDIKDAGYPKLISKGFGGLNGKIVA 242

Qy  1047 ALSTAYKKNMPESVYFFK 1064
         ||:|:|:||:|||||||||
Db   243 ALSIAQYKSRPESVYFFK 260
```

RESULT 5

```
ID   Q9N4S7       PRELIMINARY;     PRT;   1079 AA.
AC   Q9N4S7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y51B11A.1 PROTEIN.
GN   Y51B11A.1
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613;  PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Cotton M.;
RT   "The sequence of C. elegans cosmid Y51B11A.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTNSN.
SQ   SEQUENCE   1079 AA;  110532 MW;  8DBDE3824CF80CA1 CRC64;
```

```
Query Match             17.0%;  Score 967.5;  DB 5;  Length 1079;
Best Local Similarity   30.5%;  Pred. No. 8.8e-57;
Matches 321;  Conservative  133;  Mismatches  445;  Indels  153;  Gaps  43;
```

```
Qy   35 TTPDTSTTQHNKVSTSP-KITTAKPINPRPSLPPNSDTS--KETSLTTVNKETTVETKETT 91
         : |||| |:|:: || |  ||:  :||  |  |:: :  |  :  ||:|| ::|||
Db  102 TAPETTSTEAPSSSTTPVQTTTTTTAPETTSTEPPSSSSTTPVQTTTTTTAPETTSTEAPSSS 161

Qy   92 TTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPPPKAETTTKGP--ALTTPK 149
         ||:|:|  |:|  || :  |:|  |:| |  ||:  :|  || |||||||| :  || |||
Db  162 TTPVQTT-----TTTAPETTSTEPPSSS----TSPVQTTTTTTAPETTSTEPPSSSSTTPV 211

Qy  150 EPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT 209
         : |||   | | || || ||| |||||||||||||||||||||||||||||||| |||
Db  212 QTTTTTTAPETTSTEPPSSSSTTPVQTTTTTTAPETTSTESPSSSTTPVQTTTTTTAPE---TT 268

Qy  210 PKEPAPTTPKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPKEPA 269
         |||||||||||||||||| |||||||||| || ||:|||||||||||| || ||:::|
Db  269 STEP-PSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEP-PSSSTTPV 326

Qy  270 PTTPKEPAP--TTPKEPAPTAPKKPEPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTT 328
         |||||||||  |||||||| ::  | |||||||||| ||:|| ||||||| :|
Db  327 QTTTTTTAPETTSTEPPSSSS-----TTPVQTTTTT-----APETTRTEPPSSSSTTPVQNTT 376

Qy  329 KSAP--TTTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEPA 386
         |  |  || ||||||| |||| || ||:|||:|||| || |||||||||||||||||||
Db  377 TTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSST--TPVQTTTTTTAPETTSTEP- 433

Qy  387 PTTPKEPAPTTTKKPAPTAPKEP--APTTPKETAPTTPKKLTPTTPEKLAPTTPKEKAPT 444
         |||||||| |||:|| |  ||:||  :| |||| ||:|||:  | ||| |||| ||:||||
Db  434 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEP----PSSSTTPVQT 489

Qy  445 TPEELAPTTPEEP--TPTTPEEPAPTTPKEPAPTTPKEAAAPNTPKEPAP--TTPKEPAP 499
         |  | |||| ||:  | || ::  : ||| | |||| |   | ||||| |:  |||:|||
Db  490 TTTAPETTSTEPPSSSTTPVQTTTTTT----APETTSTESPSSSTTPVQTTTTTTAPETTS 545

Qy  500 TTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKEPTSTTSDKP-APTTPKGT 558
         ||| | ||||  |  ||| |||||| || ||:     |||||::|:  |  ||||||:
Db  546 TEPPSSS-TTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPV 604

Qy  559 APTTPKEPAPTTPKEP--APTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 616
         ||||||| | ||||||  |||||  |  |||  ||||| || ||:  |||| ::|  |
Db  605 TTTTAPE---TTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTST 661

Qy  617 DKP-APTTPKETAPTTPKKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPD 675
         ||  | ||||| |||| || |||  | :| :|  || ||| |||:|:|:||
Db  662 EPPSSSSTTPVQTTTTTTAPETTSTEP----PSSSNTPVQTTTTTTAPETTSTSPV 717
```

```
Qy    676 ESTPELSAEPIPKALENSPKEPGVPTTN--KTPAATKPEMTTTAKDXTTERDLRTT----P 729
              :||       :      |:||  | :    |   ||  || | |    |   |    :|:
Db    718 QTTTTAPETT------STEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAP 771

Qy    730 ETTTAAPKMTKETATTTEKTTESKIITATTTQVTSTTTQDTTPFKITTTLKTTTLAPKVTTT 789
              |:||  :       ||  :  |  |   |      |:   | |:  |||| | |: | ::
Db    772 ETTSTEPPSSSNTPVQTTTTTAPETTSTEPPSSSTTPVQTT--TITAPETTSTEPPSSSN 829

Qy    790 KKTITTTEIMNKPEETA-KPKDRAINSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRPK 848
              | |||   :  |||   |              |  |      |:|  ||||    |||
Db    830 TPVQTTT--TTAPETTSTEPPSSSTTSPVQTTTTTAPETTSTEPPSSSTTPVQTTTITAPE 887

Qy    849 TTTPTPRKMTSTMPELNPTSRIIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEG-ETPH 907
              ||  |     ||  |      |  |      || |     |      |:    |  :  |
Db    888 TTSTEPPSSSTTPVQTTTTTAPET---TSTEP-PSSSTTPVQTTTTTAPETTSTEPPS 941

Qy    908 MLLRPHVFMPEVTP---DMDYLPRV-------PNQGIIINPMLSDETN------ 945
               : |    |||     | | |            | |     |  | |
Db    942 SSTTPVQTTTTTVPDCDCSLSYIDRVVYPFTTEWWENKRDIIIQSYDSPRRTAFSFVSTFD 1001

Qy    946 ----ICNGKPVDGLITTLRNGTLVAFRGHYFWMLS------PFSPPSPARRITEVW 990
                  ||     |  ||          |   |  |       | |   |    | :
Db   1002 IGTCTAILCIYTSEGGISNL-NATL-----FIGLSDCGSSIDLPF-------Y 1041

Qy    991 GIPSPIDTVPTRCNEGKGIFFFFKDSQYWRFTN 1022
              : |     |||    ||| |||||          :
Db   1042 NVPTGLE-IWPEINCEGNK------WSYNN 1064
```

RESULT   6

Q91751

ID   Q91751;   PRELIMINARY;   PRT;   1049 AA.
AC   Q91751;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;

RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003495; AAG22353.1; -.
DR   EMBL; AE003056l; CG5228.
DR   FlyBase; FBgn0030561; CG5228.
SQ   SEQUENCE 1049 AA; 954DD629E7619671 CRC64;

Query Match   16.4%; Score 936.5; DB 5; Length 1049;
Best Local Similarity   34.6%;
Matches 371; Conservative   45; Mismatches 439; Indels 217; Gaps 54;

```
Qy      2 KDNKKNRIKKKKPTPKPKPPVVDEAGSGLDNGDFKVTTPDTS---TTQHNKVSTPKITTTAKP 58
                                                       ::       ::       |:|
Db     30 KPNEIGTTAKPTTLKP------------------------TEGTSAKPTTLKP 71

Qy     59 -----INPRPSLPPNSDTPSKETSL------TVNKETTVETKETTTTNKQT--STDGKEKTT 106
                |:|  :     |   ::|       :|:|::  |:|   :  | :|  |
Db     72 TEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG----T 127

Qy    107 SAKET--QSIEKTSAK--DLAPTSKVLAKPT------PKAETTTKGPALTTPKEPRPTT 155
              |||:     ||| |    : |  |  ||          | ::| |  :||||  ::|
Db    128 SAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLK 187

Qy    156 PKE-------PASTTPKEPT--PTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTPKEPAP 207
              |        | :||||:    ||  || ||||| ||||:|| ||||||| |:|||| ||
Db    188 PTEGTSAKPTTLKPTEGTTAKPTTLK--PTEGTSAKPTTLKP 243

Qy    208 TTPKEPAPTTTKEPAPTTTKSAPTTPKE--PAPTTPKPAPTTPKEPAPTTPKEPP---P 262
              || |    :|| ||||  |  |||| |    |||| | ||||| | ||||||||     |
Db    244 TTAK---PTTLKPTEGTTAKPTTLNPTEGTSAKPTTLKPTEGTTAKPTTLNPTEGTSAKP 300

Qy    263 TTPKEPAPTTKEPAPTTPKEPAPTAPKKKPAPTTPKEPAPTTPK------EPAPTTTKEPSP 317
              || |    || || || || | |  |   ||||| |  ||||       | ||||  |:|
Db    301 TTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG----T 360

Qy    318 TTPKEPAPTT-------TKSAPTTPKEPAPTTTKSA---PTTPKEPSPTTTKEPAPTTPKE 368
              || |         |  ||| |||| | || | |     ||||| | :|| | ||||||
Db    361 TTAK---PTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLK 416

Qy    369 PAPTTPKKPAPTTTKEPAPTTPKEPAPTTTKKPA---PTAPKEPAPTTPKETAPTTP--PK 423
              | |||| | ||||  | |||||  |||| ||       | ||| |||||| |||||   |
Db    417 --PTTLKPTEGTTAK---PTTLK-PTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTT 470

Qy    424 KLTPT------TPEKLAPTTPEKPAPTTPEELAPT---TPEEPTPTTPEEPAPTTPKAAAPN 476
               |||       || | ||||  | ||||   ||     || | ||| |||||||| |   |
Db    471 TLKPTEGTTAKPTTLKPTKGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPT 530

Qy    477 TPKEPAPTTPKEPAPTTPKEPAPTTLKPTEGTTAKPTTPKGTA------PTTLKEPAPTTP 526
              | || ||||| | |||||| |  || | ||| ||||| | |       |||||| |||||
Db    531 KGTTAKPTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTA 587

Qy    527 K----KPA------PKELAPT-----TTKEPTTTSDKPA--------PTTPKGT 558
                   |        | | |||      | ||   |  |        |||||
Db    588 KPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPT 647

Qy    559 APTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPK--KPAPKELAPTTTKGPTST 614
              ||||| | |||| | ||||   || ||||| || ||    | || | || |  | ||
Db    648 EGTTAK---PTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLNPTEG 701

Qy    615 TSDKPAPTTPKE--TAPTTPKEPAPTTPKKPA----PTTPETPPPTTSEVSTPP-TTKPT 668
              || | |||| |   |||| | | |||||        ||||  | ||||  |:|| || ||
Db    702 TSAKPTTLKPTEGTTAKPTTLKPTEGTRTTAKPTTLKPTEGTSAKPTTLKPTEGTRTTAKPT 761
```

```
Qy  669  TIHKSPDESTPELSAEPTP---------------KALENSPKEPGV--PPTKT--PAATK  709
              |:|| |:||                   :|| || ||||    |     |||
Db  762  TL--KPTEGT---SAKPTLKPTEGTTAKPTTLKPTEGTSAQPTTLK               816

Qy  710  PEMTTTAKDKTTERDLRTT----------PETTTAAPKMTKETATT------TEKTTES   752
              |||| |                     | ||   |    ||        |||:||
Db  817  PTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTTAK              876

Qy  753  KITATTTQVTSTTQDTTTPFKITTTLKTTTLLAPKVTTTKK--TITTTEIMNKPEETAKPKD  810
              |||| :|   |||| || | ||||  |    :|||   || |::   |:|  || || |
Db  877  PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLKPTEGTSAKPTTLKPTE   936

Qy  811  RATNSKRATTPKPQKPKAPK---KPT--STKKPKTMPRVRKPKTTPT---PRKMTSTMP-   861
              | || :  |:| :| ||    |||   :| ||  |   ||||| |     |:|| ||
Db  937 -GTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGPSAKPTTLKPTERTSAQPT   995

Qy  862  ELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEG-ETPHMLLRP          912
              :| ||    |  :|| | :|  |    ||        |  | :|
Db  996  TLKPTERTSAQ--PTTLKPTEGTTAKPTTLKPTKGTSGQOANNFETKRKERRP         1045
```

RESULT 7

```
ID   O76894
AC   O76894;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:56G7.1 PROTEIN
GN   EG:56G7.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003421; AAF45644.1; -.
DR   EMBL; AL031028; CAA19845.2; -.
DR   FlyBase; FBgn0025390; EG:56G7.1.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF01607; Chitin_bind_2; 2.
DR   Pfam; PF01217; P_rich_extensn; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00494; ChtBD2; 2.
SQ   SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;
```

```
Query Match        15.2%;  Score 865.5;  DB 5;  Length 1795;
Best Local Similarity  28.5%;  Pred. No. 9.2e-50;  Indels 115;  Gaps 28;
Matches   276;  Conservative  129;  Mismatches  447;
```

```
Qy    6  KNRTKKKPTPKPPVVDEAGSGLDNGDEKVTTPDTSTTTQHNKVSTSPKITTAKPINPRPSL   65
              ||| ||                    :||          |||    ||:
Db  431  KNDVTEAPEIKSPL-----KGLHLSENIVILPETTTT--TTTTTKPVVLICPTISPPDTT  483

Qy   66  PPNSDT-----------SKEISLITVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSI  114
              |||:                 ||   :|| |||  ||| |  ::  |  ||| |:| :
Db  484  PKPSTTTAVTKSTPKISSTEQHSTTTAKTTTTTKRPTTVTEKTSSATEKPRTTVVTTTQK  543

Qy  115  EKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKS  174
              | ||   || |          ||  ||  |||       | |  | |  ||    |:
Db  544  RSTTTHNTSPDTRKTTIRSTTLSPKTTTPTPSTTTPSTTTPST-----TTPSTTTPST---  594

Qy  175  APTTPKEPAPTTTKSAPTTPKEPAPTTPK-EPAPTTPKEPAPTTKEPAPTTTKSAPTTP   233
              || || |   |    |||  |    |:   |  |  |  ||   ||    |   ||
Db  595  --TTPSTTTPPSTTTTVKVSTHRERTTSQKTTTASTTTKK----TTTTSPKTTKTTDIPTST 648

Qy  234  KEPAPTTPKKPAPTTPKPEPAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTTAPKKPAP  293
              || |  |  |  ||   || ||  |  | :||  || ||  ||   |  | | |   ||
Db  649  TSKLLSTTTQKTTTTTHKFTAATTTSTEKPKTTTEKTSTVSTTRTKKSTESSPKPTSSTGKPT 708

Qy  294  TTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTKSAPTTTKEPAPTTPKEPAPTTKSAPTTPKEP 353
              |||  || |  |  ||     | |  ||  |||     |||   || |  | | | ||| |||
Db  709  TTPKPSTRTTPTTPTTKVTTTTQIITTTPLRSSTETTSTQPPTTTTPQPTTTTTTLTVTPKTS   768

Qy  354  SPTTTKEPAPTTPKEPAPTTPK----------KPAPTTPKEPAP--------TTPKE    392
               ||| || ||   |  |||          || ||  || ||        |||
Db  769  TTTTTTEKPITISSPKPTTTTTOKTTSTAPNTTKVAITTTQKETTTPPTQSTTSTTIFTRKTTTNN 828

Qy  393  PAPTTTKKPAPTTAPKEPAPTTPK-ETAPTTPKKLTPTTPKEKLAPTTPEKP-------AP   443
              | |||  |  |  | | | |||  | |||  |  | ||||  |  |||
Db  829  PEPTSTEKPITISTTPKPSTTTPKTSTVASSTEKTTISSPKPTTEKSTENPTTNSVKTSAL  888

Qy  444  TTPEELAPTTPEEPPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPT--TPKEPAPTT   501
              ||   || ||   ||||   | ||||   | ||  || |||  ||     ||| |||
Db  889  TSSTQRATSTTSEPTKTT-ONITITTTPKPTTLKTSTQEATTSTQKVSVSTVTITTKKATESS 947

Qy  502  PKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTTKEPT--STTSDKPAPTTPKGTA  559
              |  || || || ||| | | || || | | |  ||    |||     | ||| || ||
Db  948  PLTTLLSTEEPNTTPKPLRTTTPTTTISVSVTATTRITTTTTISESSTETTSTQKPKSTPTSTTT 1007

Qy  560  PTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP-TTTKGPTSTTSDK   618
              |||  || |  |  ||     | || | | || || | | |  ||  |||  | ||| |
Db 1008  RTTPKVTTVIVSTONPTTTTSKTSTVTI----TTP-NPSPSTQRPTTTTRQPSITAST    1061

Qy  619  PAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTT-TKE---------   666
```

```
Qy   1062  TSIGTRIPTTINPQNSTSSTDLTTVTRPCPDPDDSTSSDKNTNTACTQELQQVNLLELQS  1121
             |||:  | :|:      :|:      |::|       ::|:     |:|
Db   1062  PTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTPTPAATKPEMTTAKD            718
                                                            (Db 667)

Db   1122  PQKQEQFTHTRTHTALTGSRNTLGGQEVPDYMDDAPSSAEBAESGCQATTAKAPTMSTLAAA  1181

Qy    719  KTTER---DLRTTPETTTAAPKMTKETATTTESKITATTTQVTSTTTQDTTPFKIT        775
Db   1182  HLLQKLFHIISTTPPSREHAP--TQRPSSQPSSQRSR-GVTIAQMARHINLATSKPFIAH   1238

Qy    776  TLKTT----TTLAPKVTTTKKTITTTEIMNKPEETTAKPKDRATNSKATTPKPQKPTKA-PK  830
Db   1239  SLRLSIQQLASTQKRSIPPKTLVTHNTTKEPED-SEYVDSETSEQYTEDNEVLDKTQPR    1297

Qy    831  KPTTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVE  890
Db   1298  AMSSTTVAAVLPAVPSTTTEREPQK-TSSSP--SPT----KATSSTTTQPIETTTGDLEY   1350

Qy    891  VNPKSED  897
Db   1351  DSSGSSD  1357
```

RESULT   8
ID   O96449
AC   O96449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RACE 1-11;
RA   Goernhardt B.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061185; AAC72308.1; -.
SQ   SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;
                              PRELIMINARY;   PRT;   1489 AA.

```
Query Match       355;  Conservative   46;  Mismatches 475;  Indels 222;  Gaps 49;
Best Local Similarity  15.0%;   Score 853;   DB 10;   Length 1489;
Matches  355;   Conservative 32.3%;   Pred. No. 5.3e-49;

Qy      9  TKKKPKPKPPVVDEAGSGLQNGDFKVTTPDFSTT----QHNKVSTSPKITTAKPINPR      62

Qy    338  TPYAPTEKPVDVEETTYVTEESTYAPTKSETNAPTERMHYAHIEKPCDTEVTMYAPTEET  397

Qy     63  PSLP-----PNSDT---SKETSLTVNKETTVETKETTT---TNKQTSTD            100

Qy    398  TYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEETTYAPTEETTYAPIEKTTVAP   457

Qy    101  GKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKEPTPTTPKEPA   160

Qy    458  TEETTYAPTEETPYEPTEETTYAPTEETTYAPT---EETTYASTEETTYAPTEETTYAPA   514

Qy    161  STTPKEPTPPTTIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT--TTKEPAPTTPKEPAPT  217

Qy    515  EETTYEPTEET-TYAPTEETTYAPTEETTYAPAEETTYAPTEETTYAPTEETTYAPTEET   573

Qy    218  TKKEPAPTTPKSAPT----------TPKEPAPTTPKKRAPTTPKEPAPTTPKEPAPTTPKE  267

Qy    574  TYAPTEETT-YAPTEETMYAPIEETTYAPTEETTYAPAEETTPYEPTEETTYAPTEETTYA  632

Qy    268  P---------------AP----------TTKEPAPTTPKEP-----APTPARKKRAPTTPKEPAPTTP  304

Qy    633  PTEETTYASTEETTYAPTEETTYAPAEETTPYEPTEETTYAPTEETTYAPTEE          692
```

```
Qy    305  KEPAPT--TTKEPSPTTPKEP---------APT-TTKSAPT------------TTPKEP-------A   339
Db    693  TTYAPTEETTYAPTEETPYEPTEETTYAPTEETTYAPTEETMYAPIEETTYA        752

Qy    340  PT-TTKSAPT--TPKEPSPTTPKEP---------APTTPKEPAPTTPKEPAPTTPKEPAPT--   388
Db    753  PTEATTYAPTEETPYAPTEETTYEPTEETTYAPTEETTYAPTEETTYAPTEE         812

Qy    389  TPKEPAPTTPKKPAPTAPTKEPAPTTPKETAPTTPKKLIPTTPEKLAPTTPEKPAPTTPEE  448
Db    813  TPYEPTEETTYAPTEETTYVTPTEETTYAPTEETTYAPTEETTYAPTEETT          870

Qy    449  LAPT------TPEEPTPTTPKEP-----APT---------TPKAAAPNTPKEPAPTTPKEP-   489
Db    871  YAPTEEEPYEPTEETTYAPTEETTYASTEETTYAPAEETTYAPTEETPYEPT         930

Qy    490  ----APTTPKEPAPTTPKETAPTTPKGTAPT--TLKEPAPTTPKPAPK-ELAPT--TT     539
Db    931  EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETTYAPTEETTYAPMEETTYE 990

Qy    540  KEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP----------APTTPKGTAPT-------   585
Db    991  YAPTEEMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAST         1050

Qy    586  ---------TLKEPAPTTPKKP-----APKE---LAPT--TTKGPTSTTSDKPAPTTP    624
Db   1051  EETTYAPTEETTYAPAEETTYAPTEETPYEPTEETTYAPTEETTYAPTEET          1110

Qy    625  KETAPTTPKEPAPTTPKKPAPTTPEPPPTSEVSTPTTTKEPT--TIHKSPDESTPELS   682
Db   1111  YAPAEEPYEPYEPTEETTYAPTEETTYAPTEETMYAPIEETTY-APTEATTYAP       1169

Qy    683  AEPTPKA-LENSPKEPG----VPTTKTPAATKPEMTTAKDKTTERDLRTTP---ETT    732
Db   1170  TEETTYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEET         1229

Qy    733  TAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTT--------PFKITT---LKT   779
Db   1230  TYAP-----TEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDETTYGPTEE  1284

Qy    780  TTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAT--NSKATTPKP-QKPFKAPKKPT---  833
Db   1285  TTYAPTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYE  1344

Qy    834  --------STKKP-KTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP-N   881
Db   1345  PAEESTTVSTEKPCNTEEFTDEPTDEPTDE--PSDEPTDEPTDLPTDEPSTPCDN      1402

Qy    882  QTPNSKLVEVNPKSEDAG  899
Db   1403  QGINGIGVENKVRYNNAG  1420
```

RESULT   9
ID   Q9SPM0
AC   Q9SPM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN.
DE   PEX2.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=POLLEN;
RA   Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA   Hohorst D., Gao M., Showalter A., Bedinger P.A.;
                              PRELIMINARY;   PRT;   1315 AA.

RT   "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159297; AAD55980.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR001998; Xylose_isom.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS01172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR   SMART; SM00370; LRR; 3.
SQ   SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match
Best Local Similarity   25.0%; Score 808.5; DB 10; Length 1315;
Matches 255; Conservative 88; Pred. No. 4.4e-46;

Qy   13 PTP-----------------------KPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPR 62
          ||| ||:| ||| ||| :| ||: :|: | || ||| |:|
Db   412 PTPGGPPSSSVPGKPPSVPG-----------KPAAPAPMPTPHTPPDVSP-------EPL-PE 455

Qy   63 PSLPP---------------NSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQ 112
          |:|||| || :| | |: | |||:| |||: : | | | : |:|
Db   456 PSPVPAPAPMRMPTLRSPPADEYIPT----PEVPAKSPPGTSPPASRGAPPLQAQPPAAS 511

Qy   113 SIEKTSAKDLAPTSKV-----------------LAKPTPKAETTTKGPALTTPKEPTPTTPKE 158
          || | ||:: | | | ||||||| |||||||
Db   512 SPPATPVKSSPPPAAVVLPPPAKTPSPPEAPVASPPEAPVSSPQPQVKSSPPPAPVASPP 571

Qy   159 PASTTPKEPTPTPTTIKSAPTTPKEPAPTTTKSAPTTP--KEPAP----------TITKEPAPTTP 210
          ||| || | | | || || |: | || | :| ::|| ||| |||| |||||
Db   572 PPMKSP--PPPARVASPPPLMKSPPPPAPVASPPPLKSSPPPVLWLSTPSVKSPPPVP 629

Qy   211 KEPAPTTTKEPAPTTTKSAPTTPKE----PAP---TTP----KKPAPTTPKEPAPTTPKE- 259
          ||||||||||||||| |||||||||| ||| ||| |:||||||||:|||| ||
Db   630 VASPPPVKSPPPLAPVSSPPSPPPVKLPPLPAPGKSTPPPEEEKPTPPTTPVKSSPPPEKSL 689

Qy   260 PTTPTTPKEPAPTTPKEPAPTTTKSPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTT 319
          ||||||||||:| :|| |||||| :: || |||||||||||||:| |||||||||
Db   690 PPPPLITTSPPQEKPTPSTPSKPPPSPVELLPPPPSKSSPPPEEPVSSPPQAPKSSSP-- 747

Qy   320 PKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTPKEPSPTTTKEPAPTTPKEPAP 379
          || | ||| || | | :|| | || | ||:||||| ||||||
Db   748 ---PAPVS--SPPPLKSSPPPVPESSPPTPTPKSSPPLAPVSSPPQVEKTSPPPAPVSSPP 802

Qy   380 TTPKEPAPTTKKPAPTTAPKEPAPTTPKKLTPPTTPEKLAPTTPE 439
          :| ||| || || | | |||||||||| | | |:||||
Db   803 PTPKSSPPLAPVSSPPQVEKTSPPPAPVSSPPPVEKTSPPPAPV 862

Qy   440 KPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAANTPKEPAPTTPKEPAPTTPKEPAP 499
          |||||| | | | | || | || || |||| | |:||||||||| |||||
Db   863 SSPPLEPKSPSPSVSSPPTVKSSPPMTPKSSSPPAHVSSPPPEAEKSSPP 922

Qy   500 TTPKETAPTTPKGTA--------------------PTTLKEPAPTTPKKPAPKELAPTTPK 549
          |||||||||| | ||| ||||:| |||| | |
Db   923 LAPISSPPSEPKSPSSPPWVEKTSPPPATVSSPPPTPKSSPPP--- 979

Qy   549 KPAPT---------TPKGTAPTTP---------KEPAPTTPKEPAPTTPK--GTAPTTLKEPAPT 593
          ||||| || | |||| ||||||||| |:|||||||
Db   980 PPAPVSSPPPKPLPPPAPVSSPPPVVKSSPPPTPVSSPPPTPKPLPPPPTTPVSSPPPT 1039

Qy   594 TPKKPAPKELAPTTPKGPTSTSTSDKPAPTT----PKETAPTTPK------EPAPTTPKKPAPTT 647
          |||| | | | :: | | ||| | |||||| ||||||||| | ||
Db   1040 XKPLPPP---APVSSPPVVKSSPPPAPVSLPPPPVKCCPPPTL 1096

Qy   648 PETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA 707
          |||||| | | || |||:| | |||| ||||| | || |||| |||| |||||||||
Db   1097 VSSPPPAPKSLPPPIPVSSP-----------PPEVKSSPPPTPVSSPPPAPKSSSPPPTVS 1146

Qy   708 TKPEMTTAKDKTTERDLRTTP-----ETTTAAPKMTKETATTTEKTTESKITATTTQVT 762
          ||| | | |||| ||||| | | |||| | ||:| | :|| | |:| |
Db   1147 SPPP----------ELKSSPPPAPVSSPPSAPKSSPPPAPVSLPPPEVKSSPPPAPIS 1194

Query Match
Best Local Similarity   14.2%; Score 808.5; DB 10; Length 1315;
Matches 255; Conservative 88; Pred. No. 4.4e-46; Indels 219; Gaps 33;

RESULT   10
Q20007

Qy   763 STTTQDTTPFKITTLKTTTLAPKVTTTKKTTITTTEIMNKPEETAKPKDRATNSKATTPKP 822
          :| | |:
Db   1195 SPPPPAKSP--------------------PPPA------------PMSSLPPVKSPPPPAPVSSP 1228

Qy   823 QKPFKAPKKPTSTKKPKTMPRVRKPKTTPTTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQ 882
          | | ||:||| |:| | | || || | |:|
Db   1229 PPPMKSPPPPAPISSPPPAP--VKPPSLPPPAPVSSPPPAV------------TSAP-- 1271

Qy   883 TPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD 942
          :| | : | | | | || | | |:: | |:||
Db   1272 -----------PKKEDSTA------------------PPAEALPPPSFNDILLPPIMAN 1302

Qy   943 E 943
          |
Db   1303 K 1303

Query Match
Best Local Similarity   14.1%; Score 802; DB 5; Length 1274;
Matches 290; Conservative 102; Pred. No. 1.1e-45; Indels 230; Gaps 56;

Qy   6 KNRTKK-KP-----TP-----KPPVVDEAGSGLDNGD---FKVTTPDTSTTQHNKVSTS 50
          || ||| || | ||| :: |||||||||||||||||||

```
ID   Q9LIE8;      PRELIMINARY;   PRT;   1480 AA.
AC   Q9LIE8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP001306; BAB03062.1; -.
DR   InterPro; IPR002965; P.rich.extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   1480 AA;   D1AC0C79F155E7732 CRC64;

Query Match              14.0%;
Best Local Similarity    29.6%;                 Score 800.5;  DB 10;  Length 1480;
Matches  307;  Conservative   54;  Mismatches 505;  Indels 171;  Gaps 45;
                                                Pred. No. 1.7e-45;
```

RESULT 11
Q9LIE8

```
Qy     10  KKKPTKPKPPVVDEAGSGLDNGDPFKVTTPDTSTTQHNKVSTS----PKITTAKPINPRSL  65
           | :| :||||                        ::  | :  |   |:||||||| ||||
Db     66  PNSDTSKETSLTVNKETTIVEFKETTTTNKQTSTDGKEKTISAKETQSIEKTSAKDLAPTI 125

Qy    117  KPRPHPKPPVNK-------PHPHPKPTKPHPHPKPPTKHHPHPKPPTIKPPPKPPSV  167
           || :||                ||| |   || |      | ||     |:|||
Db    168  KPPPSTPK----------------------PPTTNPPPSTPQQPTHKPPPCTPTPPVVASPPMATTP 211
```

(Protein sequence alignment — Query (Qy) vs Database (Db) pairwise alignment continuing across the page.)

```
Qy    622 TPPIATPPIAKPPVATPPTATPPIATSPVAKPVATPPTKTPPPAKPPVAIPPIATPPVA 681
Db    540 KEPISTTSDKPAPTTPKGPTAP--TTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKP 598
Qy    682 KPPVAT---PPTATPPTATPPVTTPIATPPVVTPPIATPPVTPIAKPPTTI--PPTATPPVA 736
Db    599 SPPPPVKSPPPAPVGSPPPPEEKSPPPPAPVASPPPPVKSP--PPPTLVASPPPPVKSPP 599
Qy    599 APKELAPTTPKGPTST--TSDKPAPTTPKETAPT-TPKEPAPTTPKKPAPTTP-----E 649
Db    184 PTTTKSAPTTP--KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKSAPTTP-KEPAPTT 240
Qy    737 MPPIATPPTAKPPIATPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVA 796
Db    600 PPAPVASPPPPVKSPPPPTTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPP 659
Qy    650 TPPTTSEVSTPTTTKEPTTIH--KSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA 707
Db    241 PKKPAPTTPKEPAPTT------PKE--PTPTTPKEPAPTTPKEPAPTTPKEPAPTTP 291
Qy    797 TPPIAKPPVAIPPTTAPPTATPPVAKPPVAIPPTTKTPPPTAKPPILIPPISKPPVATPPAA 856
Db    660 PAKSTTPPPEEYPTPPTSVKSSPPPPEKSLPPPTLIPSPPPQEKPTTPPSSTPSKP-PSSPEKP 718
Qy    708 TKPEMT-TTAKDKTTTERDLRTTPETTTAAPKMKETATTTEKTTESKITATTTQVTSTTT 766
Db    719 SP--PKEPVSSPPQTP---KSSPPPPAPVSSPPPTPVSSPPPALAPVSSPPPSVKSSP--- 768
Qy    767 QDTTPFKITTLKTTTLAPKVTTTKKT--IITTEIMNKPEETAKPKDRATNSKATTKP 822
Db    769 PPAPLSSPPPAPQVKSSPPPPVQVSSP-PPAPKSSPPLAPVSSPPPQVEKTSPPPAPLSSPP 827
Qy    823 QKPTKAPKKPISTKKPKTMPRVRKPKTTPTPRKMTSTMPELNP-------TSRIAF--- 871
Db    828 LAPKSSPPPSPHVVSSPPPVKSSPPPAPVSSPPLTPKPASPPAHVSSPPEVKRPSTP--PA 885
Qy    872 -----AMLQTTTRPNQTPNSKL--VEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMD 924
Db    886 PTTVISP-PSEPKKSSPPPTPVSLPPPIVKSSPPPAMVSSP-PMTPKSSPPPVVVSSPPPT 943
Qy    915 PAKPPVAIPPIATPPVAKPPVATPPTATPPIATSPIATPPVVT---PPTATSPVATPPIA 971
Db    944 VKSSPPPAPVSSPPPATPKSSPPPAPVNLPPPEVKSSPPPTPVSSSPPPA---PKSSPPPAP 1000
Qy    925 YLPRVPNQGIIINPMLS 941
Db   1001 MSSPPPPEVKSPPPPAPVSSPPPPVKSPPPPAPVSSSPPPV 1057
Qy   1026 VLPPIAKPPVETSPTATPPTATPPVAIPP-------VVKPPVAIPPITKPPV 1070
Db   1058 KSPPPPAPVSSPPPPVKSPPPPAPVS--SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS 1115
Qy   1071 ATPPVTNPPTAMPPIVT 1087
Db   1116 SPPPAP-----VKPPSLP-PPAPVSSPPPVVTPAPPKKEEQSLPPPRAES 1158
```

RESULT 12

```
Q41805

ID   Q41805         PRELIMINARY;      PRT;   1188 AA.
AC   Q41805;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z34465; CAA84230.1; -.
DR   Mendel; 14346; Zeama.2368;14346.
DR   InterPro; IPR001611; LRR.
DR   Pfam; PF00560; LRR; 3.
DR   SMART; SM00370; LRR; 4.
KW   Signal.
FT   SIGNAL        1      27      POTENTIAL.
SQ   SEQUENCE   1188 AA;  120980 MW;  2C77C7F8D7130149 CRC64;
```

RESULT 13

```
P70670

ID   P70670         PRELIMINARY;      PRT;   2187 AA.
AC   P70670;
DT   01-FEB-1997 (TREMBLrel. 02, Created)
DT   01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE   MUSCLE-SPECIFIC FORM GP220).
GN   NACA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96312450; PubMed=8698236;
RA   Yotov W.V., St-Arnaud R.;
RT   "Differential splicing-in of a proline-rich exon converts alphaNAC
RT   into a muscle-specific transcription factor.";
RL   Genes Dev. 10:1763-1772(1996).
DR   EMBL; U48364; AAB18734.1; -.
DR   EMBL; U48363; AAB18732.1; -.
DR   MGD; MGI:106095; Naca.
DR   InterPro; IPR002715; NAC.
```

```
Qy   564 -KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTTKGPTSTTSDKPAPT 622
Db  1001 MSSPPPPEVKSPPPPAPVSSPPPPVKSPPPPAPVSSP---PPVKSPPPPAPVSSSPPPV 1057
Qy   623 -TPKETAPTTPKKPAPTTPETPPPPTTSEVSTPTTTKEPTTIHKSPDESTPEL 681
Db  1058 KSPPPPAPVSSPPPPVKSPPPPAPVS--SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS 1115
Qy   682 SAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTTERDLRTTPET 731
Db  1116 SPPPAP-----VKPPSLP-PPAPVSSPPPVVTPAPPKKEEQSLPPPRAES 1158
```

```
Qy   511 APVG-----------------SPPPPVKTTSPPAPIG--SPSPPPPVSVV 541
Db   129 -----LAKPTPKAETTTKGPALTTPKEPAPTTPKEPTTPKEPAPTTPKEPA 183
Qy   542 SPPPPVKSPPPAPVGSPPPPEEKSPPPPAPVASPPPPVKSP--PPPTLVASPPPPVKSPP 599
Db   184 PTTTKSAPTTP--KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKSAPTTP-KEPAPTT 240
Qy   600 PPAPVASPPPPVKSPPPPTTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPP 659
Db   241 PKKPAPTTPKEPAPTT------PKE--PTPTTPKEPAPTTPKEPAPTTPKEPAPTTP 291
```

Query Match   13.8%;
Best Local Similarity   29.98%;
Matches 230;   Conservative 59;   Mismatches 355;

```
Qy    13 PTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRSL----PPN 68
Db   460 PTPHSPPAD--------DYVPPTPPVGKSPPATSPSPQVPPAASTPPPSLVKLSPPQ 510
Qy    69 SDTSKEFSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV 128
```

Score 784;   DB 10;   Length 1188;
Pred. No. 1.7e-44;
Indels 122;   Gaps 33;

```
DR   InterPro; IPR003037; TS-N.
DR   Pfam; PF01849; NAC; 1.
DR   Pfam; PF02094; TS-N; 1.
SQ   SEQUENCE 2187 AA; 220599 MW; 003646AA864DEBFD CRC64;

Query Match       13.6%;   Score 776;   DB 11;   Length 2187;
Best Local Similarity   27.7%;   Pred. No. 1.1e-43;
Matches 288;   Conservative 122;   Mismatches 389;   Indels 240;   Gaps 49;

Qy    34 VTTPDFSTTQHNKVSTSPKITTAKPINPRSLPPNSDTSKE------TSLTVNKETT 84
            |||:|  |||:||:  :|:  :: | :  ::
Db   806 VQSPKVDPIMSDVTPTSPKKTSATAV-PK------DTSATLSLKSVPAVTSLSPPKAPV 857

Qy    85 VETKETTTNKQTSTDGKETTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPA 144
            |: :  :||  :|| : ::  |  |::
Db   858 APSNEATIVPEIPTSLKNALAAATPKETLATSIPKVTSPS-----PQKIPKSVSLKGAPA 913

Qy   145 LTT----------PTPTTPKEPASTT------PKEPTPTT 171
            ::|               |||      |||  :
Db   914 MTSKRATEIAASKDVSPSQPPKEVPLLQHVPPTSPPKSPVSDTLSGCALTSPPPKGP-PAT 972

Qy   172 IKSAPTTPKE-PAPTTTKSAPTTPK-------EPAPTTTKEPAPTTPKEPAPTTTKEP 221
            |:| |  |  | |:   |  ||       | |:  |||| :
Db   973 LAETPTYPKKSPKPAASKKTPATPSPEGVTAVPLEIPPCSKPKAPKTAAPKESSATSSSKR 1032

Qy   222 APTT--TKSAPT---------------TPKEPAPTTPKK-PAPTTPKEPAPTTP 257
            | || ||| |               ||||  ||  ||  :
Db  1033 APKTAVSKEIPSKGVTAVPLEISLPLKETSKSASPKSPKIAGPKE---TPP 1089

Qy   258 KEPTPTTPKEPAPTTKEPAPTTKEPAPTTPKKLTPTTPKE--PAPTTPEELAPTTPE 311
            |  | ||| |  ::  ||  ||  | |:  |
Db  1090 GGVTAVPPEISLPPKETPQNATPNESLAASSQKRSPKITSVPKETPPGGVTAMPLEIPSAP 1149

Qy   312 TKEPSPTTPK-------EPAPTTKSAPTTTKEPAPTTT-KSAPTTPK------EPSPTT 357
            |:| | || |        | |:  ||   :: |||||  || |||
Db  1150 QKAPKTAVPKQIPPEDAVTILAGSPLSPKKASKTAAPKEAPATPSVGIAVSGEISPSP 1209

Qy   358 TKEPAPTTPKEPAPTTPKKPAPTT--PKE--------PAPTTKEPAP-------T 396
            |:| |  ||||| |  |  | |    |||        | |||| | |       |
Db  1210 KKTSKTAAPKENSATLPFKSRSPKTAAPKETPATISSEGVTAVPSEISPSPPTPASKGVPVT 1269

Qy   397 TTTKPAPTTAPKEPAPTTPKKLTPTTPKEKLAPTTTPEK-PAPTTPEELAPTTPE 455
            | |: |  | ||||| ||   ||  |    | |||| |  | |||||| | ||||
Db  1270 LTPKGAPNALAE-SPASPKVPKTAAPEETSTTP------SPQKIPKVAGPKEASAAPPS 1322

Qy   456 EPTPTT-------PEEPAPTTPKAAAPNTPKEPAPTTPKE-PAPTTPKEPAPTTPKETA 506
            |  ||          | || ||   | ||| |  ||||  | |||| | |||||  |
Db  1323 KKTPKTAVPKETSAPSEGVTAVPLEIPPSPRKAPKTAAPKEFPAPS--PEGATTAPVQIP 1380

Qy   507 PTTPKGTAPTTLKEPAPTTP------------KKPAPKELAPTTTKEPTS-- 544
            | |||  | ||| |  |||             || |  | |||| ||  :
Db  1381 PSPRKGSKKAGSKE-TPTTPSPEGVTAAPLEIPISSKKTSKMASPKETLVTPSSKKLSQT 1439

Qy   545 ----TTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKKPA 599
                |||  | ||  |  | |||  |  ||||  | |||  | ||| | |  |||| | |
Db  1440 VGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-A 1496

Qy   600 PKELAPTTTKGPTSTSDKPAPTTPKETAPTTPKEPAPTTPKE-PAPTTPETP----PPT 654
            ||  |  ||: |:||   | | || |  |  | | |  ||   | ||||||     |||
Db  1497 PKTAAPPSER-VTTVPPEKPA-TPQKASGTTASKVPVPAETQEVAVSSREPVTPAVPPV 1554

Qy   655 TSEVSTPPTTK-----EPTTIHKSPDESTPELSAEPTPKALENSPKE-PGVPTTKTPAA 707
            :|: |   ||      | |:  |:  :||  :  | |||| |||||  |  || | |
Db  1555 KNPSSHKKTSKTIELKEPAPALLPPSPTKSPKIPSSKKAPRT--SAPKEFPASPSIK-PVT 1611

Qy   708 TKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQ 767
            |  |  ::  | | :  | : ||   ||
Db  1612 T--SLAQTAPPSLQKAPSTTLPKENLAAPAV---- 1665

Qy   768 DTTPFKITLKTTLAPKVTTTKKITTTTELMNKPEETARPKDRATNSKATTPKPQKPTK 827
            :||| |  |:||  ||  ||:||  ||| :|:| || | || | | ||| ||||| ||
Db  1666 QTAPKEATTIPSCKKAAATETPIEISTAPSLEGAPKETSE---TSVSKVLMSSP----- 1716
```

RESULT 14
Q9VEL9

```
Qy   828 APKKPTSTKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSK 887
            ||||: ||: || :: |  |  |:
Db  1717 -PKKASSKRASTLP------------AHTLPSLKEASVLS----------PTATSSGK 1752

Qy   888 LVEVNPKSEDAGGAEGETP 906
            : :| ||:  |  |  ||
Db  1753 DSHISPVS-DACSTGTTTP 1770
```

ID   Q9VEL9   PRELIMINARY;   PRT;   2112 AA.
AC   Q9VEL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG4090 PROTEIN.
GN   CG4090.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li J., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003716; AAF54402.1; -.
DR   FlyBase; FBgn0038492; CG4090.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR005611; EGF-like.
DR   Pfam; PF01607; Chitin_bind_2; 1.
DR   SMART; SM00494; ChtBD2; 1.
DR   PROSITE; PS00022; EGF_1; 1.

```
SQ SEQUENCE 2112 AA; 219547 MW; B91018E5551A5D36 CRC64;
Query Match
Best Local Similarity   13.0%;
Matches   281; Conservative   98;    Score 739; DB 5; Length 2112;
                                     Pred. No. 3e-41; Indels 316; Gaps 44;
                                     Mismatches 339;

Qy    4 NKKNRIKKKPPPKPPVVDEAGSGLDNGDFKVTTPDJTSTTQHNKVSTSPKITTAKPINPRP   63
        : :: ||  ||         ||
Db 1023 NHPDQIQTKPLKCKKVVSQGGSSSN------STSNSSSSSSNNSGSSSNSGGSSSS-----  1070

Qy   64 SLPPNSDTPSKEFSLTVNKETTV---ETKETTTTNKQTS----TDGKEKTTSAKETQSI--  114
        || |||              ::                   ::: |:   ||
Db 1071 SSSSNSGGSSSNTGSSSNSGASSSSGGGSSNSGGSSSGGSNSSGGNQSTSSTSSSSSSSS   1130

Qy  115 -----EKTSAKDLAPTSKVLAKPTPKAETTTKGP--------------------------  143
          :: || |:  :    |||
Db 1131 NNNNQGSSSSSSSSSSSSSKPNPSETCKVNGQFIGDRSDCAKFYRCVDNDRGGFNMVPF    1190

Qy  144 ----------ALTTPKEPTTPTPKEPASTTPKEPTTPTTIKSAPTT                178
                  ||| |:||  ::|||  | :||: ::|||
Db 1191 SCGPGTVWDAQMACNIHAWAVKECGGIAPPTTTTSPTTSR--PTTASTSRPSDQTSTSRPTG 1249

Qy  179 PKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTT----------------------  217
         ||||                 ||||
Db 1250 P------PTTARPVTARPTTSSPTTASSSQTTTSPVTQAPNTDGKCRSEGPMADPNNCSKFYR 1305

Qy  218 --------TKEP---------APT----------                            248
                 |||          |||
Db 1306 CVRNNKGGFTSIPFQCGAGTVWDQDLQTCNHHFNNCSTGTESTTPKPPC--EPATNGTTA   1363

Qy  249 PKEPAPTTPKEPTTTKEPAPTTTKEPAPTAPKKPAPTTPKEPAPTTPKEPA           308
         ||||  |||      ||||             ||| |||
Db 1364 TSTSSTTTP--PPTTTDLPPTSTTGLP-PTTTTELPPT------TTTDLPPTTTTRLPP    1413

Qy  309 PTTTIKEPSPTTPKEPAPTTTKSAPTTTTKSAPTTTPK---EPSPTTTPKEPAP        363
         ||||       ||||              ||| |
Db 1414 TTTTSLPPTTTTGLPPTTTTGAQPTTTTLSSETETSTVVTTSPESTTQPPSTTTMKPLPAG  1473

Qy  364 ---------------TTPK-------------------EPAP---TT                373
                        ||||                   ||||
Db 1474 TECTGEGYMADPEDCRKYYRCINAGASYRKYNFTCPKGTGGWNEEVQTCDYVENIPRCSKL   1533

Qy  374 PKKPAPTTP-----KEPAPTTPKE-PAPTTTKKPAPTAPKEPAPTTPKEPAPTTPKKPAPTTPKE 428
         |||                |||      |||
Db 1534 PAEPITTTTFSEESKDPGSTTPQSTDEPTTVTKPI-TKPTE-EPSTEKPOKPTIQYPEKPT    1591

Qy  429 TPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKE    488
         | ::            ||               ||
Db 1592 TTEE--PEKPQKPI-TTEYPOKPTTPEEPIPEKPOK                           1647

Qy  489 PAPTI-PKEPAPTTPKETA-------PTTPK--GTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
         |||  :        |:         |||
Db 1648 PTTTEYPQKPTTEEPTTTSIPGYNPTTSVPGVNPTTTPIPVETTTSTPGYK---PTTTG     1704

Qy  541 EPTSTTTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKK-   597
           ||              |||               |||
Db 1705 EPITTTTT-LPSTTTDAIQEPTTSKKPEPTTTTESPESSTPEGSVTTLQPEPQPNYNCSSE   1763

Qy  598 ---PAPKELA-------PTTIKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTP  648
            |||               |:|                ||| |||
Db 1764 GFFPDPEDCSRYRCVDAAKNGKYQYYAFKCGKGTVWDTSTETCNYADQVSGN------     1816

Qy  649 ETPPPTTSEVSTPTTTPKEPTTTHKSPDESTPELSAEPTPKALENSPKEPGVPTIKPAAT   708
          |||                |||
Db 1817 ------CSSGQTTTPGTTTEPGTTTESTTSSGKP---ETTSKAPEN-----            1857

Qy  709 KPEMTTTPAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQD 768
         || |||        || ||||
Db 1858 -PETTTTT-----SSPETTTT---VASETTTTSGT------TTTTATPETTTKP         1895

Qy  769 TTPFKITTLKTTTLAPKVTTTFKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKA 828
         : || ||| || :|||   :|| : ||| |
Db 1896 PKP-------ETTTIAGEEISTSKSPTTTE---------SPAPSTNTSA              1928

Qy  829 PKKPISTKKPKTMP                                                842
         | |
Db 1929 P--------CPETGP                                               1935
```

RESULT 15
Q9XDH2

ID   Q9XDH2;   PRELIMINARY;   PRT;   763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH MUCIN HOMOLOG.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang Y., Moreno C., Singh M.;
RT   "cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   InterPro; IPR002565; P_rich_extensn.
DR   InterPro; IPR003802; Pistil_extensin.
DR   PRINTS; PR01222; ATROPHIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE 763 AA; 39168EC45A5916F8 CRC64;

```
Query Match
Best Local Similarity   12.5%;
Matches   248; Conservative   39;    Score 715; DB 2; Length 763;
                                     Pred. No. 4.5e-40; Indels 232; Gaps 39;
                                     Mismatches 361;

Qy  132 PTPKAETTTKGPALTTPKEPTTPTTPKEPASTTPKEPTTPTTIKSAPTTPKEPAPTTIKSAP 191
        ||
Db    3 PVP-------APRALALPLPPAPPAPAEPKSKPPEPPAP---------PAPPCWMLVSAAP   46

Qy  192 TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTTTPKKPAPTTPKE 251
        |||                                                     |||
Db   47 PCP-PAP----PAPPKPKSKAPPFPVPPAPPAPARELAPPLP--PAP-----PEAPRE     90

Qy  252 PAPTTPKEPTTPTT-----PKEPAPTTKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK 305
        |||              ||                    ||
Db   91 SRPALPPCPPPPPVVIPDPPEPAAPPVPPAPNSPPPPFPPPAPPKVPAAPPVP--PVPNSPP 148

Qy  306 EP--APTTTKEPSPTTPKEPAPTTTKEPAPTTTKSAPTTTKSAPTTTKEPSPTTTKEPAP  363
        |:                |||
Db  149 FPPPPPAALNPPAP-----PAPPLANSPPLPPAPPPAGT--PPAAAWPPVPAAPKSKPA    201

Qy  364 TTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAP----TTTKKPAPTAPAKPEPAPTTPKETAP 419
         |||             |||               ::|
Db  202 SPPRPPAP-----PMPATPMEFPPLPPVPPDPISKETPAPPAPPIPPAPVPIPPVPPLP     256

Qy  420 TTPKKLTPTTPEKLAP----TTTPEKPAPTTPEELAPTTTPEEPTPTTPEEPAPTT        469
        ||                  |:||
Db  257 PVPNKIPPAPP---APPVAVAVAAVLVAPCPLLPPLPNNHPPAPPAAPVPGVPLAPLPNSHP   313

Qy  470 ---PKAAAPNTPKEPAP--------------                               483
           ||||
Db  314 PAPPSAPVPGVPLAPLPISGRPVSWKGSFTTLSTFCCRVCSGEVLAGALNPSRSPSPL     373

Qy  484 --TTPKEPAPTTPKEPAPTTPKETA----PTTPKGTA---PTTLKEPAPTTP-KKPAPKE   533
          |||             |:||              |||
Db  374 TTTTPALPAPIPPLPPLPLPINTAVPPIPPLPPVVTALAPPLPPLAPLPISPGVPPAP--   431
```

```
Qy   534  LAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPA------PTTPKEPAPTTPKGTAPTTL  587
              || :| :||  ||| ||| |||| ||: ||     | |:   |:
Db   432  --PIPGKPWTTPPLAPAPPEPK-TVPVLPPGPSCPPSEKPNPPAPPEPEPKSSPALPP  488

Qy   588  KEPAPTTP---KKPAPKELAPTTTKGP-TSTTSDKPAPTTPKET---APTTPKEPAPTTP  640
              |||: | :    | :    |||:  |||  ||: |||| ||:      ||| |||| |
Db   489  APPAPSMPSAVRVPPSPPIPPAPPAAPRASMPALPAPPSSPPATRLCPPLPPSPPAPNSP  548

Qy   641  KKPAPTTPETPPPPTTPSEVSTPTTTKEPTTIHKSP-----DESTPELSAEPTPKA--LENSP  694
              ||| | ||| ||| ||| :  |||||  |||           :: |  :   |||    |||
Db   549  --PAPPAPPTPPKLLS--ANPPCPPVPPAPNRPPAPPAPPELPAPDDPPTPPVANSP  604

Qy   695  KEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKI  754
              |:| :| |    :| :         |   |:|                ||||
Db   605  --PAPPAPPAPPSALPFVNPPA-------------PPTPAAPK--------SRPAL  637

Qy   755  TATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATN  814
              |  :  |||                              | | |
Db   638  PAAPPAPPAPPVRATTP--------------------------PPAPPAPPAPN  665

Qy   815  SKATTPKPQKP----TKAPKKPTSTKKKPKTMPRVRKPKTTPTPRKMTSTM-----PELNP  865
              | : | | ||    |  |  |  |   |:          |  |  |       | |:|
Db   666  SMALPPAPPDPPIPLLATPPAPPAPPLPMSPPAPPLPPRAAPDDPPAPPLTINQPPSPPLAP  725

Qy   866  TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET  905
              |                |: |:|
Db   726  VPGAPLAPLPINGRPVFARKNSLI--------GSSSGDT  756
```

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: April 26, 2002, 16:22:33 ; Search time 153.5 Seconds
(without alignments)
965.302 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1212
Perfect score: 5416
Sequence: 1 VKDNKNRTKKKPTPKPPVV..........SPIDTVFTRCNCEGKTFFFK 1013

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5416	100.0	1404	4	Q92954	homo sapien
2	5407	99.8	1404	4	Q94185	homo sapien
3	2294	42.4	1054	11	Q9BX49	homo sapien
4	967	17.9	1079	5	Q9mj99	mus musculu
5	936.5	17.3	1049	5	Q9N4S7	caenorhabdi
6	865.5	16.0	1795	5	Q9I7S1	drosophila
7	853	15.7	1489	10	Q91s1	drosophila
8	829.5	15.3	401	6	O76894	drosophila
9	808.5	14.9	1315	10	O96449	phytophthor
10	802	14.8	1274	10	O77765	bos taurus
11	800.5	14.8	1480	10	Q9SPM0	zea mays (m
12	784	14.5	1480	10	Q20007	caenorhabdi
13	776	14.3	1188	10	Q9LIE8	arabidopsis
14	739	13.6	2187	11	Q41805	zea mays (m
15	715	13.2	2112	5	P70670	mus musculu
16	707.5	13.1	763	2	Q9VEL9	drosophila
17	671	12.4	555	10	Q9XDH2	mycobacteri
18	667.5	12.3	1151	13	Q9JLT1	rattus norv
19	667.5	12.3	5085	11	Q9JKS6	rattus norv

ALIGNMENTS

RESULT 1

20	660.5	12.2	6677	5	Q9N435	caenorhabdi
21	659.5	12.1	1229	5	Q94185	caenorhabdi
22	657.5	12.1	2284	5	Q9vpg1	drosophila
23	653.5	12.1	1514	5	Q9guw7	leishmania
24	647	11.9	4833	11	Q90qx6	mus musculu
25	647	11.9	5038	11	Q9qyx7	mus musculu
26	643	11.9	2089	4	Q14676	homo sapien
27	633	11.7	489	10	Q41707	vigna ungui
28	632.5	11.7	7962	4	Q10465	homo sapien
29	632	11.7	761	10	Q9zq10	arabidopsis
30	632	11.6	6632	5	O17362	caenorhabdi
31	629	11.6	3507	5	O23587	caenorhabdi
32	628.5	11.6	1612	5	Q9vyq2	drosophila
33	626.5	11.6	990	13	Q91803	xenopus lae
34	625	11.5	6642	5	O01761	caenorhabdi
35	624.5	11.5	2768	5	Q9vc00	drosophila
36	622.5	11.5	839	2	Q9rx57	deinococcus
37	622	11.5	3570	4	Q99552	homo sapien
38	620.5	11.5	971	5	Q9xvs4	caenorhabdi
39	615	11.4	2344	5	Q9n3y8	caenorhabdi
40	607.5	11.2	801	5	O23635	caenorhabdi
41	607.5	11.2	924	12	Q99307	epstein-bar
42	605	11.2	379	5	Q27929	drosophila
43	600.5	11.1	1893	5	Q9nkc9	drosophila
44	598.5	11.1	409	10	Q9sbm1	volvox cart
45	577	10.7	956	10	Q9lj64	arabidopsis

ID   Q92954   PRELIMINARY;   PRT;   1404 AA.
AC   Q92954;
DT   01-FEB-1997 (TREMBLrel. 02, Created)
DT   01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   MEGAKARYOCYTE STIMULATING FACTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA   FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA   Jacobs K., Turner K.;
RL   (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL   Mosher D.F. (eds.);
RL   Biology of vitronectins and their receptors., pp.45-52,
RL   Elsevier Science Publishers B.V. (1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Blood 78:279-279(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
DR   EMBL; U70136; AAB09089.1; -.
DR   InterPro; IPR005585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   InterPro; IPR002400; GF_cysknot.
DR   Pfam; PF00045; hemopexin; 2.
DR   Pfam; PF00086; Somatomedin_B; 2.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR00438; GFCYSKNOT.

```
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
SMART; SM00120; HX; 2.
SMART; SM00201; SO; 2.
SQ   SEQUENCE   1404 AA;  151090 MW;  AABD7AD19B35F4F6 CRC64;

Query Match             100.0%;  Score 5416;  DB 4;  Length 1404;
Best Local Similarity   100.0%;  Pred. No. 0;  Mismatches   0;  Indels   0;  Gaps   0;
Matches 1013;  Conservative  0;

Qy      1 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTIPDTSTTQHNKVSTSPKITTAKPIN 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTIPDTSTTQHNKVSTSPKITTAKPIN 259

Qy     61 PRPSLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 PRPSLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 319

Qy    121 DLAPTSKVLAKPTPKAETTKGPALTTPKEPPTTTPKEPPTTTPKEPPTTTIKSAPTTPK 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 DLAPTSKVLAKPTPKAETTKGPALTTPKEPPTTTPKEPPTTTPKEPPTTTIKSAPTTPK 379

Qy    181 EPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTIKSAPTTPKEPAPTT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 EPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTIKSAPTTPKEPAPTT 439

Qy    241 PKKPAPTTPKEPAPTTTPKEPAPTTPTPTTPKKEPAPTTPAPKKPAPTTPKEPA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    440 PKKPAPTTPKEPAPTTTPKEPAPTTPTPTTPKKEPAPTTTPAPKKPAPTTPKEPA 499

Qy    301 PTTPKEPAPTTTTKEPSPTTPKEPAPTTTTKSAPTTTKSAPTTPKEPSPTTTKE 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    500 PTTPKEPAPTTTTKEPSPTTPKEPAPTTTKSAPTTTKSAPTTPKEPSPTTTKE 559

Qy    361 PAPTTPKEPAPTTTTKKPAPTTKPAPTTPKRAPTTTKEPAPTTTKEPAPTTTPKEPA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 PAPTTPKEPAPTTTTKKPAPTTKPAPTTPKRAPTTTKEPAPTTTKEPAPTTPKETAPT 619

Qy    421 TPKKLTPTTPKEKLAPTTPEKLAPTTPEELAPTTPEEPTTTPEEPAPTTTPKAAAPNTPKE 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 TPKKLTPTTPEKLAPTTPEELAPTTPEELAPTTPEEPTTTPEEPAPTTTPKAAAPNTPKE 679

Qy    481 PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKGTAPTTLKEPAPTTPKKELAPTTTK 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 PAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKELAPTTK 739

Qy    541 EPTSTTISDKPAPTTPKGTAPTTLKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 EPISTTISDKPAPTTPKGTAPTTLKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 799

Qy    601 KELAPTTTKGPTSTTISDKPAPTTPKETAPTTPKKPAPTTPETTPPPTTSEVST 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    800 KELAPTTTKGPTSTTISDKPAPTTPKETAPTTPKKPAPTTPETTPPPTTSEVST 859

Qy    661 PTTTKEPTTIHKSPDESTPELSAEPTKRALENSPKEPGVPTTKPAATKPEMTTTAKDKT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    860 PTTTKEPTTIHKSPDESTPELSAEPTKALENSPKEPGVPTTKIPAATKPEMTTTAKDKT 919

Qy    721 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    920 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT 979

Qy    781 TLAPKVTTTKKTTITTTEIMNKPEETAKPDRATNSKATTPKPQKPTKAPKKPTSTKKKPT 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    980 TLAPKVTTTKKTTITTTEIMNKPEETAKPDRATNSKATTPKPQKPTKAPKKPTSTKKKPT 1039

Qy    841 MPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTTNSKLVEVNRKSEDAGG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1040 MPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTTNSKLVEVNRKSEDAGG 1099

Qy    901 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1100 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN 1159

Qy    961 GTLVAFRGHYFWMLSPFSPPSSPARRITEVWGIPSPIDTVFIRCNCEGKTFFPK 1013
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1160 GTLVAFRGHYFWMLSPFSPPSSPARRITEVWGIPSPIDTVFIRCNCEGKTFFFK 1212
```

```
RESULT 2
Q9BX49
ID   Q9BX49;            PRELIMINARY;          PRT;   1404 AA.
AC   Q9BX49;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).
GN   BG174L6.2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133553; CAC36090.1; -.
SQ   SEQUENCE   1404 AA;  151076 MW;  782A11746B3FDEE5 CRC64;

Query Match             99.8%;  Score 5407;  DB 4;  Length 1404;
Best Local Similarity   99.8%;  Pred. No. 0;  Mismatches   2;  Indels   0;  Gaps   0;
Matches 1011;  Conservative  0;

Qy      1 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTIPDTSTTQHNKVSTSPKITTAKPIN 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTIPDTSTTQHNKVSTSPKITTAKPIN 259

Qy     61 PRPSLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 PRPSLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 319

Qy    121 DLAPTSKVLAKPTPKAETTKGPALTTPKEPPTTTPKEPPTTTPKEPPTTTIKSAPTTPK 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 DLAPTSKVLAKPTPKAETTKGPALTTPKEPPTTTPKEPPTTTPKEPPTTTIKSAPTTPK 379

Qy    181 EPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTIKSAPTTPKEPAPTT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 EPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTIKSAPTTPKEPAPTT 439

Qy    241 PKKPAPTTPKEPAPTTTPKEPAPTTPTPTTPKKEPAPTTPAPKKPAPTTPKEPA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    440 PKKPAPTTPKEPAPTTTPKEPAPTTPTPTTPKKEPAPTTTPAPKKPAPTTPKEPA 499

Qy    301 PTTPKEPAPTTTTKEPSPTTPKEPAPTTTTKSAPTTTKSAPTTPKEPSPTTTKE 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    500 PTTPKEPAPTTTTKEPSPTTPKEPAPTTTKSAPTTTKSAPTTPKEPSPTTTKE 559

Qy    361 PAPTTPKEPAPTTTTKKPAPTTKPAPTTPKRAPTTTKEPAPTTTKEPAPTTTPKEPA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 PAPTTPKEPAPTTTTKKPAPTTKPAPTTPKRAPTTTKEPAPTTTKEPAPTTPKETAPT 619

Qy    421 TPKKLTPTTPKEKLAPTTPEELAPTTPEELAPTTPEEPTTTPEEPAPTTTPKGTAPTTLKEPAPTTPKKELAPTTTK 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 TPKKLTPTTPEKLAPTTPEELAPTTPEELAPTTPEEPTTTPEEPAPTTTPKAAAPNTPKE 679

Qy    481 PAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 PAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTK 739

Qy    541 EPTSTTISDKPAPTTPKGTAPTTPKGTAPTTLKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 PAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTK 739

Qy    541 EPTSTTCDKPAPTTPKGTAPTTPKGTAPTTLKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 EPTSTTCDKPAPTTPKGTAPTTPKGTAPTTLKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 799
```

```
Qy   601 KELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKKPAPTTPEPTPEPPTTSEVST 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   800 KELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKKPAPTTPEPTPEPPTTSEVST 859

Qy   661 PTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKT 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   860 PTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKT 919

Qy   721 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   920 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT 979

Qy   781 TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTTPKPQKPTKAPKKPTSTKKPKT 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   980 TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTTPKPQKPTKAPKKPTSTKKPKT 1039

Qy   841 MPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRRNQTPNSKLVEVNPKSEDAGG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1040 MPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRRNQTPNSKLVEVNPKSEDAGG 1099

Qy   901 AEGETPHMLLRPHVEMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1100 AEGETPHMLLRPHVEMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN 1159

Qy   961 GTLVAFRGHYFWMLSPFSFPPSPARRITEVWGIPSPIDTVFTRCNCEGKTFFFK 1013
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1160 GTLVAFRGHYFWMLSPFSFPPSPARRITEVWGIPSPIDTVFTRCNCEGKTFFFK 1212
```

```
RESULT   3

ID   Q9JM99;      PRELIMINARY;      PRT;   1054 AA.
AC   Q9JM99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE   PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN   PRG4
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ikegawa S., Nakamura Y.;
RT   "a novel mouse gene gene highly homologous to a human gene encoding
RT   megakaryocyte stimulating factor precursor and cartilage superficial
RT   zone protein.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034730; BAA92310.1; -.
DR   MGD; MGI:1891344; Prg4.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR002955; P_rich_extensn.
DR   InterPro; IPR001212; Somatomedin_B.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
DR   PROSITE; PS00024; HEMOPEXIN; 2.
DR   PROSITE; PS00201; HEMOPEXIN_UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ   SEQUENCE   1054 AA;   115991 MW;   4FC64BFA42283235 CRC64;
```

```
Query Match             42.4%;  Score 2294;  DB 11;  Length 1054;
Best Local Similarity   47.9%;  Pred. No. 1.7e-141;
Matches  491;  Conservative  35;  Mismatches 130;  Indels 368;  Gaps 20;

Qy     1 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVT--TPDTSTTQHNKVSTSPKITTAKP 58
         ||||||.||:||.:|||||||||||||||||:||.....||:|:|||||||.|:||||||
Db   195 VKDNKKNTPKKKPNPEPPAVDEAGSGLDNGEFKLIPPPDPDPPTTPHSKVATSPKTTAAKP 254
```

```
RESULT   4

ID   Q9N4S7;      PRELIMINARY;      PRT;   1079 AA.
AC   Q9N4S7;
```

```
Qy    59 INPRPSLPPNSDTSKETISLTVNKETTVETKETTTINKQTSTDGKEKTTSAKETQSIEKTS 118
         .:||||.||||.|||..||.:.|||.|..|||||||||||..|:||.|||:||||      
Db   255 VTPKPSLAPNSETSKEASLASNKETTVETKETTATNKQSSA--SKKKTSVKETRSAEKTS 313

Qy   119 AKDLAPTSKVLAKPTPKAETTTKGPALITPKEPTPTTPKEPASTTPKEPPTTIKSAPTT 178
         .:.|..||.|.:..|.|..|||||||||||||||||||||.||||||||||||:|||||
Db   314 DKDV----------------------------------------------EPTSTTPK 331

Qy   179 PKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAP 238
Db   332 TKKPV-TTTKESKFLP----LPQEPEPTTAKEPPTTTKKPEPEPTTPKEPEPTTTKEPEP 376

Qy   239 TTPKKPAPTTPKEPAPTTPKEPAPTTTPKGTAPTTLKEPAPTTPKKP 298
Db   377 TTPKEPEPTTTPKEPEPTTTPKEPEPTTTPKEPGPTTPKEPEPTTKE 436

Qy   299 PAPTTPKEPSPTTTPKKPAPTTTPKEPSPTTTPKEPSPTTT 358
Db   437 PEPTTTKEPESTTRKEPEPTTPKEP 474

Qy   359 KEPAPTTTPKKPAPTTPKEPAPTTPKEPAPTTTPKKPAPTAPKEPAPTTPKETA 418
Db   475 KEPEPTTPKEPEPTTTPKEPEPTTPKEPEPTTPKEPEPT 512

Qy   419 PTTPKKLPPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAANTNP 478
Db   513 --------------------------------------------------------TP 514

Qy   479 KEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT 538
Db   515 KEPEPTTPKEPEPTTPKEP--------------------KEPEPTTPKKP-- 549

Qy   539 TKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKP 598
Db   550 --------EPTTPKEPVPTTPKEPTTP------KEPEPTTPKEP 581

Qy   599 APKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKRPAPTTPETPPPPTTSEV 658
Db   582 EP--------:-TTPKKEPEPTTTPKEPEPTTTPKEPEPTTPKKPEPTT 617

Qy   659 STPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKIPAAATKPEMTTTAKD 718
Db   618 ------------------------------------------------------------ 617

Qy   719 KTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLK 778
Db   618 ------TSPKT---------------------------------------------TTLK 626

Qy   779 TTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAINSKAT--------TPKP-QKPTKAP 829
Db   627 ATTLAPKVTAPAE-------EIQNKPEETTPASEDSDDSKTTLKPQKPTKKPTKAP 681

Qy   830 KKPTSTKKPKTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRRNQTPNSKLV 889
Db   682 KKPTSTKKPKI--PKTRKPKTTPAPLKTTSATPELNTTP--LEVMLPTTTIPKQTPNPETA 738

Qy   890 EVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNG 949
Db   739 EVNPDHEDAGGGEGEKP-LIPGPVLFPTAIPGTDLLAGRLNRGININPMSDETNLCNG 797

Qy   950 KPVDGLTTLRNGTLVAFRGHYFWMLSPFSFPPSPARRITEVWGIPSPIDTVFTRCNCEGKT 1009
Db   798 KPVDGLTTLRNGTLVAFRGHYFWMLNPFRPPSPPRRITEVWGIPSPIDTVFTRCNCEGKT 857

Qy  1010 FFFK 1013
Db   858 FFFK 861
```

DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y51B11A.1 PROTEIN.
GN   Y51B11A.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Cotton M.;
RA   "The sequence of C. elegans cosmid Y51B11A.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match   17.9%; Score 967; DB 5; Length 1079;
Best Local Similarity   32.3%; Pred. No. 3.2e-55;
Matches 294; Conservative 114; Mismatches 377; Indels 124; Gaps 33;

Qy   35 TTPEDTSTTQHNKVSTSP-KITTAKPINPRRSLPPNSDTS--KETSLTVNKETTVETKETT 91
             | ||| ::::::|| | | :: ||| :| | : :| ::|| | : ::
Db   102 TAPETTSTEPAPSSSTTPVQTTTTTTAPETTTSTEPPSSSTSPVQTTTTTTTAPETTSTEAPSSS 161

Qy   92 TTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKRAETTTKGP--ALTTPK 149
             | ||| :| || | | | | | ||| | | :| |||
Db   162 TTPVQTT-----TTTAPETTSTEPPSSS---TSPVQTTTTTTTAPETTSTEPPSSSSTTPV 211

Qy   150 EPTPTTPKEPAPSTTPKEPPTTIKSAPTTPKEPAPTTTKEPAPTTTKEPAPTT 209
             |||| || || | |: | :| |: |||| || ||| :||| |||
Db   212 QTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTESPSSSSTTPVQTTTTTTAPE---TT 268

Qy   162 TTPVQTT----------------TTTAPETTSTEPPSSS----TSPVQTTTTTTTAPETTSTEPPSSSSTTPV 211

Qy   210 PKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTTKKPAPTTPKEPAPTTPKEPTTPTTPKEPA 269
             || | |||| ||| |: |||| || ||| :| ||| || ||| || |||
Db   269 STEP-PSSSTTPVQTTTTTTAPETTSTEPP-SSSSTTPV 326

Qy   270 PTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKEP-------------APTTTKEP 315
             |||| || || || || | |||| ||| |||||
Db   327 QTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTE 386

Qy   316 SPTTPKEPAPTTTKSAP---TTTKEPAPTTTTKEPSPTTTKEP-APTTPKEPAPT 372
             ||| ||| ||| :| : | |:|||| ||: ||| || |||
Db   387 PPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPP-SSSSTTPV 446

Qy   373 TPKKPAPTTPKEPAPTTTKKPAPTTAPKED--APTTPKE----TAP------ 419
             || |||| || | || :|||| | | :| ||||| |||
Db   447 T----APETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSSTTPVQTTTTTTTAPETTSTEP 502

Qy   420 ----TTPKKLTPTTPEKLAPTTPEKPAP---TTPEELAPTTPEEPPTTPEE---- 468
             :|| : || |: ||| ||| :| || ||||
Db   503 PSSSSTTPVQTTTTT----APETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSSTTPVQT 558

Qy   469 TPKKAAAPNTPKEP---APTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 526
             ||| :: | |: :|| ||| | || | |:| : ||| ||| ||||
Db   559 TTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPP-SSSS-TTPVQTTTTTTAPETTSTEP 617

RESULT 5
Q9I7S1
ID   Q9I7S1   PRELIMINARY;   PRT;   1049 AA.
AC   Q9I7S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

Qy   527 KKPAPKELAPTTTKEPTSTTSDKP-APTTPKGTAPTTPKEPAPTTPKEP--APTTPKGTA 583
             ||| ||| || |: :| || ::|: | |||| | | ||| | || |||| |:
Db   618 PSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPE---TTTSTEPPSSSTTPVQTT 674

Qy   584 PTTLKEPAPTTPKPAPKELAPTTTKGPTSTTSDKP-APTTPKETAPTTPKKP 642
             ||| | | ||| ||| | ||| |:| :| || ::| :| ||| |||||
Db   675 TTTAPETTSTEPPSSSNTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTTAPETTSTEP 732

Qy   643 PAPTTPETTPPTTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPT---PRALENSPKEPGV 699
             || | || || ||| | || || :|| ::||: | :| :| :||||| ||
Db   733 --PSSSTTPVQTTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSSNTP---V 786

Qy   700 PTTKTPAAATKPEMTTTAKDKTTERDLRTT----PETTTAAPKMTKETATTTEKTTTESKIT 755
             | | |: | :|| || :| |: ::   :||| |:|| |: | ||:| :| :| |
Db   787 QTTTT---TAPETTSTEPPSSSTSPVQTTTTTTAPETTSTEPPSSSNTPVQTTTTTTTAPETT 843

Qy   756 AFTTQVTSTTTQDTTTPFKIITILKTTITLLAPKVTTTK----KTIITTEIMNKPEETA-KPK 809
             | | : |: | | | | :|| | || || |||:| : || | |||
Db   844 STEPPSSSTTSPVQT-----------TTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPP 894

Qy   810 DRATNSKRATTTPKPQKPTKAPKKPTSTKKKPKTMPRVRKPKTTPPTPRKMTSTMPELNPTSRI 869
             :| | | :|| || | :| | | |: | || ||:: | |||| |:||:| |||: |
Db   895 SSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSSTTP-------- 946

Qy   870 AEAMLQTTT 878
             :|||
Db   947 ----VQTTT 951

RA Nelson D.R., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Reese M.G.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003495; AAG22353.1; -.
DR FlyBase; FBgn0030561; CG5228.
SQ SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;

Qy Query Match   17.3%; Score 936.5; DB 5; Length 1049;
    Best Local Similarity   34.6%;
Matches 371; Conservative 45; Mismatches 439; Indels 217; Gaps 54;

Qy   2 KDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTS---TTQHNKVSTSPKITTAKP 58
         |||| :: | || ||| | | | | :|||||
Db   30 KPNEIGTTAKPTTLKP------------------------TEGTSAKPTTLKPTEGTSAKPTTLKP 71

Qy   59 -----INPRPSLPPNSDTSKETSL---------TVNKETTVETKETTTTNKQT--STDGKEKTT 106
         |||| |: |:| :||| | | :: ||
Db   72 TEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG----T 127

Qy   107 SAKET--QSIEKTSAK--DLAPTSKVLAKPT-----------PKAETTLKPALTTPKEPPTT 155
         ||||| | | : || || | | | |
Db   128 SAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLK 187

Qy   156 PKE-------PASTTPKEPT----PTTIKSAPTTPKEPAPTTIKSAPTTPKEPAPTTPKEPT---P 207
         ||| | | : || | | ||| | ||| ||||||||
Db   188 PTEGTSAKPTTLKPTEGTSAKPTTLK--PTEGTSAKPTTLKPTEG 243

Qy   188 PTEGTSAKPTTLKPTEGTSAKPTTLK--PTEGTTAKPTTLKPTEG 243

Qy   208 TTPKEPAPTTTKEPAPTTPKE---PAPTTAKPTTPKEPAPTTPKEPT----P 262
         ||| || | | |||| || | | |||||| || ||| |
Db   244 TTAK---PTTLKPTEGTTAKPTTLKPTEGTTAKPTTLNPTEGTSAKP 300

Qy   244 TTAK---PTTLKPTEGTTAKPTTLKPTEGTTAKPTTLNPTEGTSAKP 300

Qy   263 TTPKEPAPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPK------EPAPTTTKEPSP 317
         ||| |||| ||||| ||| | | || ||| | |||||||
Db   301 TTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTDG 360

Qy   301 TTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTDG 360

Qy   318 TTPKEPAPTT--------TKSAPTTTKEPAPTTTKSA---PTTPKEPSPTTTKEPAPTTPKE 368
         || | || || |||||||||| | | ||| | |||||||||
Db   361 TTAK----PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAK- 416

Qy   361 TTAK----PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAK- 416

Qy   369 PAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKKPA---PIAPKEPAPTTPKETAPTT--PK 423
         |||| ||| ||||||||| | ||||| | || | ||||||||| |||| ||
Db   417 --PTTLKPTEGTTAK---PTTLK-PTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTAK 470

Qy   417 --PTTLKPTEGTTAK---PTTLK-PTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTAK 470

Qy   424 KLTPT-------TPKEKLAPTTPPKPAPTTPEELAPT---TPEEPTTTPPEEPAPTTPKAAAPN 476
         :| | |||||||| ||| || | || |||||||
Db   471 TLKPTEGTTAKPTTLKGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAK 530

Qy   471 TLKPTEGTTAKPTTLKGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAK 530

Qy   477 TTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTA---------PTTLKPEPAPTTP 526
         ||| || |||||| | ||| | | |||||||||
Db   531 KGTTAKPTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTA 587

Qy   531 KGTTAKPTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTA 587

Qy   527 K-------KPA------PKELAPT------------------TTKEPTSTTSDKPA-- 558
         : || | |||| |||| |
Db   588 KPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPT 647

Qy   559 APTTPKEPAPTTPKEPAPTTPKGTAPTTLKPEPAPTTPK----KPAPKELAPTTPKGPTST 614
         |||||||||||||||| || | ||||||| |||| ||| ||||| | ||
Db   588 KPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPT 647

Qy   648 EGTTAK--------PTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLNPTEG 701
         | |||| |||| |||| | ||||||| |||||||| | || |||||
Db   648 EGTTAK---PTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLNPTEG 701

Qy   615 TSDKPKPAPTTPKE--TAPTTPKEPAPTTPKKPA----PTTPEPPPTSEVSTPT--TTKEPT 668
Db   648 EGTTAK---PTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLNPTEG 701

Db   702 TSAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTRTTAKPTI 761

ID   O76894
AC   O76894;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:56G7.1 PROTEIN.
GN   EG:56G7.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M.,
RA   Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K.,
RA   Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C.,
RA   Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E.,
RA   Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R.,
RA   Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

PRELIMINARY;   PRT;   1795 AA.

Qy   669 TIHKSPDESTPELSAEPTP----------------------KALENSPKEPGV--PTTKT--PAATK 709
         |: | | :|| | | | ||:|| | :| |
Db   762 TL---KPTEGT---SAKPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTSAQPTTLK 816

Qy   710 PEMTTTAKDKTTERDLRTT-------------------PETTTAAPKMTKETATT------TEKTTES 752
         || ||| | | || | | || | | ||| | ||
Db   817 PTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAK 876

Qy   753 KITATTQVSTSTTTQDTTPFKITTTLKTTTTLAPKVTTTKK--TTTTEIMNKPEETAKPKD 810
         || || | | | || | : | |||| | :||: ||| ||| |||
Db   877 PTTLKPTEGTSAKPTTLKPTDGTTAKPTTLKPIDGTTAKPTTLKPTEGTSAKPTTLKPTE 936

Qy   811 RATNSKATTPKPQKPTKAPK----KPT--STKPKTMPRVRKPKTTPT---PRKMTSTMP- 861
         ||: || | | | | | | | | || || | | |
Db   937 -GTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAK 995

Qy   862 ELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEG-ETPHMLLRP 912
         | | | | | | | || | | | | |
Db   996 TLKPTEGTSAQ--PTTLKPTEGTTAKPTTLKPTGQANNPETKRKERRP 1045

RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RN   Benos P.;
RP   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003421; AAF45644.1; -.
DR   EMBL; AL031028; CAA19845.2; -.
DR   FlyBase; FBgn0025390; EG:56G7.1.
DR   InterPro; IPR0025557; Chitin_binding.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF01607; Chitin_bind_2; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00494; ChtBD2; 2.
SQ   SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match   16.08%; Score 865.5; DB 5; Length 1795;
Best Local Similarity   28.5%; Pred. No. 2e-48;
Matches 276; Conservative 129; Mismatches 447; Indels 115; Gaps 28;

```
Qy      6 KNRTKKKPPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL 65
            |||  |  | ||||     :|:|:|:||  |||:|   ||:|||:: |:::|
Db    431 KNDVTEAPEIKSPL----KGLHLSENIVILPETTTT--TTTTKPVVLTCPTISPDTT 483

Qy     66 PPNSDT--------------SKETSLTVNKETTVETKETTTNKQTSTDGKEKITSAKETQSI 114
            | |            |:|:||||||||:||||||||| ::|: |:|||  ||:||
Db    484 PKPSTTAVTKSTPKISSTEQHSTTAKTTTTKRPTTVTVEKTSSATEKPRTTVVTTTTQK 543

Qy    115 EKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPTPKEPASTTPKKIKS 174
            | ::| | :| ||||:|:|||:|||||||:||||||:|:||||:
Db    544 RSTTTHNTSPDTKTTIRSTTLSPKTTTTPSTTTPST-----TTPSTTTPST--- 594

Qy    175 APTTPKEPAPTTTKSAPTTPKEPAPTTTK--EPAPTTTPKEPAPTTTKSAPTTP 233
            |::|   || ||||||| |   |||||       ||||| |:|
Db    595 --TTPSTTTPSTTTTVKVSTHRPRTTSQKTTTASTTRK---TTTSPKTTKTTDIPTST 648

Qy    234 KEPAPTTPKKPAPTTTKEPAPTTPKEPTTPTPKEPAPTTTKEPAPTTPKEPAPTTAPKPAP 293
            | :| ||  |||| |:|||| |  |   |||| :|||||:|||| |:|
Db    649 TSKLSTTTQKTTTTTHKFTAATTSTEKPKTTTEKTSTVSTTTKKSTESSSPKPTSSTGKPT 708

Qy    294 TTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTMPKEPAPTTTKSAPTTPKEP 353
            ||||||||||||||||:|||  | ||||||| ||:||| |||||||| ||:||||||
Db    709 TTPKPSTRTPTTTKVTTTTQITTTTPLRSSTETTTSTQPPTTTTPQPTTTTTLLVTPKTS 768

Qy    354 SPTTTKEPAPTTPKEPAPTTPK-----------KRAPTTPKEPAP------------TTPKE 392
            : |||| ||||| |||||||              |||||||||              ||||
Db    769 TTTTTTTEKPITSSPKPTTTISTVASSTEKTTISSPKPTEKSTENPTTNSVKTSAL 828

Qy    393 RAPTTTKKPKPAPTTAPKEPAPTTPK-ETAPTTPKKLITPTPEKLAPTTPKEP 443
            :||||   | :| | |||||||| :|:|: |:||| |  |: |  |||
Db    829 PEPTSTEKPIITSSTTPTPKISTVASSTEKTTISSPKPTEKSTENPTTNSVKTSAL 888

Qy    444 TTPEELAPTTPEEEPTPTPKEPAPTTPKKEPAPTTPKEPAPT--STTSDKPAPTTPKGTA 501
            : |:|| | |  | |: | | ||| |  |   |||| :||||    |||::|| |:|
Db    889 TSSTQRATSTTSEPPKTT-QNITTTTPKPTTLKTSTQEATTSTQKVSTVSIVIITTKKATESS 947

Qy    502 PKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTPKEPT---TTSDKPAPTTPKKTA 559
            :  | || |: |  |:|| |||| | |     || |||||       ||::|| |:|
Db    948 PLTTLLSTEEPNTTPKLRTTTPTTISTVATTRITTITTTTISESSTETTTSTQKPKSTTPSTT 1007

Qy    560 PTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP-TTTKGPTSTTSDK 618
            ||||| || ||| || | |     |:|| |     |||  || |    |||::|| |:
Db   1008 RTTPKVTTVIVSTQNPTTTISKTSTVVTI----TTP-NPSPTQRPTTTTRQPTSITTAST 1061
```

```
Qy    619 PAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTT-TKE------ 666
            |: | |::|   | |   |:| |      |::| |||:|||:|| | |
Db   1062 TSIGTTRIPTTINPQNSTSSTDLTTVRPPCDPDDSTSDKNTNTACTQEBLQQVNLLELQS 1121

Qy    667 ------PTTIHKSPDRSTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTITAKD 718
                  |||:|  |||   |||||| |     |:|:  || |  ||||:||:|
Db   1122 PQKQEQFTHTRTHTALTGSRNTLLGGQEVPDYMDDAPSSAEAEASGQATTAKAPTTMSLLAAA 1181

Qy    719 KTTER--DLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTTPKIT 775
            |:|:|     | |: | |::|  |: || || |:|   | ||:|||:||||||||||
Db   1182 HLLQKLFHIISTTPPSREHAP--TQRPSSQPSSSQORSR-GVTIAQMARHNLATSKPFIAH 1238

Qy    776 TLKTT----TLAPKVTTTKKTTITTTEIMNKPEETTAKPKDRATNSKATTPKPQKPTKA-PK 830
            |:|:|    |:|:| |||| |:|||:|:||||||||||||: ||:||::|||: | |:|
Db   1239 SLRLSIQQLLASTQKRSIPPKTLVTHINTTKEPED-SEYYDSETSEQYTDEDNEVLDKTQPR 1297

Qy    831 KPTSTRKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVE 890
            |||:|:||| |:| |:| |||:|:| |:|| || |  |  ||| |:||: |::| |:|
Db   1298 AMSSTTVAAVLPAVPSTTTEREPQK-TSSSP--SPT----KAISSTTTQPIETTTGDLEY 1350

Qy    891 VNPKSED 897
            |:|
Db   1351 DSSGSSD 1357
```

RESULT 7

ID   096449
AC   096449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)   PRELIMINARY;   PRT;   1489 AA.
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RACE 1-11;
RA   Goernhardt B.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061185; AAC72308.1; -.
SQ   SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

Query Match   15.7%; Score 853; DB 10; Length 1489;
Best Local Similarity   32.3%; Pred. No. 1.1e-47;
Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

```
Qy      9 TKKKPPTPKPPVVDEAGSGLDNGDFKVTTPDTSTT-----QHNKVSTSPKITTAKPINPR 62
            | |||  |   |  |:|:  || ||      | |    |:|| |:|| | |:
Db    338 TPYAPTEKPYDVEETTVVTEESTYAPTKSETNAPTERMHYAHIEKPCDTEVTMYAPTEET 397

Qy     63 PSLP-----------PNSDT----SKETSLTVNKETTVETKETTT---TNKQTSTD 100
            | |            |:|:|    ||||:|: ||| |  |  |||    ||::|:|
Db    398 TYAPTEETYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPA 457

Qy    101 GKEKTTSAKETQSITEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPA 160
            | | || |:| | |:  |:| | || ||| |  |  |:|| |   | || |  |||:|
Db    458 TEETTYAPTEETYEPTEETTYAPTEETTYAPT---EETTYASTEETTYAPTEETTYAPA 514

Qy    161 STTTPKEPTPTTIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT--TTKEPAPTTPKEPAPTT 217
            | |  |:|  | |:|| |||| |||    |:|| |||| |||    ||:|| |||| |||
Db    515 EETYEPTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEET 573

Qy    218 TKEPAPTTTKSAPT----------TPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTT 267
            |:|| ||:|| |           ||| || |  | || ||||| ||:|| |||
Db    574 TYAPTEETT-YAPTEETMYAPIEETTYAPTEETTYAPAEETTYEPTEETTYAPTEETYA 632

Qy    268 P------------------AP--TTTKKEPAPTTPKEP----------APTTAPKKPAPTTPKEPAPTT 304
            |                  ||  | |:|| |||| |          ||| |:|| | || |||| |||
Db    268 AP-----------------APTTAPKKPAPTTPKEPAPTT 304
```

```
Db  633  PTEEITYASTEEITYAPTEEITYAPAEEITPYEPTEEITYAPTEEITYAPTEE  692
Qy  305  KEPAPt---TTKEPSPTTTPKEP------APT-TTKSAPT-------------  339

Db  693  TTYAPTEEITYAPAEEITPYEPTEEITYAPTEEITYAPIEEITYAPTEEITYA  752
Qy  340  PT-TTKSAPT------TPKEPSPTTTKEP-------------------------  388

Db  753  PTEATTYAPTEEITYAPTEEITYEPTGEETTYAPTEEITYAPTEEITYAPTEE  812
Qy  389  TPKEPAPTTTKKPAPTAPKEPAPTTPKEPAPTTTPKKLTPTTPKEKLAPTTPKEKPAPTTPEE  448

Db  813  TPYEPTEEITYAPTEEITPYEPTEET--TYTPTEEITYAPTEEITYAPTEEITT  870
Qy  449  LAPT--TPEEPTPTTPEEP------------TPKAAAPNTPKEPAPTTPKEP--  489

Db  871  YAPTEEITPYEPTEEITYAPTEEITYASTEEITYAPTEEITYAPAEEITPYEPT  930
Qy  490  -----APTPTKEPAPTTPKKETAPTTPKGTAPT--TLKEPAPTTPKKPAPK-ELAPT--TT  539

Db  931  EEITYAPTEEITYAPTEEITYAPTEEITYAPAEEITYAPTEEITYAPTEEITT  990
Qy  540  KEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP-----APTTPKGTAPT--  585

Db  991  YAPTEEITYAPIEEITYAPTEEITYAPTEEITYAPTEEITYAPTEEITYAST  1050
Qy  586  -------TLKEPAPTTPKKP------ARKE---LAPT--TTKGPTSTTSDKPAPTTP  624

Db  1051 EEITYAPTEEITYAPAEEITYAPTEEITYAPTEEITYAPTEEITYAPTEEITT  1110
Qy  625  KETAPTTPKEPAPTTPKKPAPTTPETPPTTSEVSTPTTTKEPT--TIIHKSPDESTPELS  682

Db  1111 YAPAEEITPYEPTEEITYAPTEEITYAPIEEITYGPTEEITY-APTEATTYAP  1169
Qy  683  AEPTPKA-LENSPKEPG-----VPTTKTPAATKPEMTTIAAKDKITERDLRTTP----ETT  732

Db  1170 TEEITPYAPTEEITYEPTGEETTYAPTEEITYAPTEEITYAPTEEITYEPTEET  1229
Qy  733  TAAPKMTKETATTTEKTTESKITATTTQVISTTTQDTT-------------  779

Db  1230 TYAP----TEEITYEPTEEITYAPTEEITYAPTEEITYAPIEEITYGPTEE  1284
Qy  780  TTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAT--NSKATTTPKP-QKPTKAPKKPT--  833

Db  1285 TTYAPTEATTYAPTEEITYAPTEEITYAPTEEITYAPTEEITYAPMEETPYE  1344
Qy  834  -------STPKKP-KTMPRVRKPKTTPPTPRKMTSTMPELNPTSRIAEAMLQTTTTRP--N  881

Db  1345 PAEESTSTVSTEKPCNTEEFTDEPTDEPTDE--PSDEPTDEPTDLPTDEPSTPCDN  1402
Qy  882  QTPNSKLVEVNPKSEDAG  899

Db  1403 QGINGIGVENKVRYNNAG  1420
```

RESULT 8
077765
ID   077765   PRELIMINARY;   PRT;   401 AA.
AC   077765;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ARTICULAR CARTILAGE;
RA   Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA   Aydelotte M.B.;
RT   "Immunodetection and Partial cDNA Sequence of the Proteoglycan,
RT   Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT   Joints.";
RL   J. Orthop. Res. 0:0-0(1998).
DR   EMBL; AF056218; AAD13404.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   Pfam; PF00045; hemopexin; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   SMART; SM00120; HX; 2.
FT   NON_TER   1
SQ   SEQUENCE   401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;

Query Match   15.3%; Score 829.5; DB 6; Length 401;
Best Local Similarity   75.4%; Pred. No. 1.1e-46;
Matches 156; Conservative 17; Mismatches 31; Indels 3; Gaps 3;

RESULT 9
Q9SPM0
ID   Q9SPM0   PRELIMINARY;   PRT;   1315 AA.
AC   Q9SPM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN.
GN   PEX2.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=POLLEN;
RA   Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA   Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT   "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159297; AAD55980.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR001999; Xylose_isom.
DR   InterPro; IPR0029965; P_rich_extensn.
DR   InterPro; IPR0035592; LRR_out.
DR   Pfam; PF05560; LRR; 3.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS01172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR   SMART; SM00370; LRR; 3.
SQ   SEQUENCE   1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

```
Qy  808  PKDRATNSKATTTPKPQKPTKAPKKPTSTKKKPKTMPRVRKPKTTPTPRK-MTSTMPELNPT  866
Db  5    PKGRATNSQVTTPKPQKPTKAPKKPTSTKKPKT-PRVRKPKPKTTPTPPSKTTTSAMPEPTPT  63

Qy  867  SRIAEAMLQTTTTRPNQTPNSKLVEVNPKSEDAGGAGEGETPHMLLRPHVFMPEVTPDMDYL  926
Db  64   S-LPEAMLQTTTRPTTPTPNSEIIDVNSENEDGDAAEGEKPHMIFRPPVLTPIVIPGTEII  122

Qy  927  PRVPNQGIIINPMLSDETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPSPSPSPARRI  986
Db  123  VRGPSQGFGINPMFSDETNLCNGRPVDGLTTLRNGTLVAFRGHYFWMLTPTPTPPPPRRI  182

Qy  987  TEVWGIPSPIDIVFTRCNCEGKTFFFK  1013
Db  183  TEVWGIPSPIDIVFTRCNCEGKTFFFK  209
```

Query Match   14.9%; Score 808.5; DB 10; Length 1315;
Best Local Similarity   25.0%; Pred. No. 7.8e-45;
Matches 255; Conservative 88; Mismatches 459; Indels 219; Gaps 33;

```
Qy  13   PTP----------KPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPR  62
```

RESULT   10
Q20007

```
Db   412  PTPGGPPSSSVPGKPPSVPG----------KPAAPAMPTPHTPPDVSP-----EPL-PE  455
Qy    63  PSLPP--------NSDISKETSLTVNKETTVETKETTTNKQTTSDGKEKTTSAKETQ    112

Db   456  PSPVPAPAPMRPTLRSPPADEYIPT-----PPVPAKSPPGTSPPASRGAPPLQAPPAAS  511
Qy   113  SIEKTSAKDLAPTSKV----------LAKPTPKAETTTKGPALTTPKEPTTTPKE      158

Db   512  SPPAPPVKSSPPPAAVLPPPAKTPSPPPAPVASSPPEAPVSSPPQPQVKSPPPAPVASPP  571
Qy   159  PASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTP-KEPAP-----TTTKEPAPTTP   210

Db   572  PPMKSP---PPPARVASPPPLMKSPPPAPVASPPQPLKSPPPPVLWLSTPSVKSPPPVP   629
Qy   211  KEPAPTTPKEPAPTTTKSAPTTPKE---PAP---TTP---KKPAPTTPKEPAPTTPKE-  259

Db   630  VASPPPPVKSPPPLAPVSSPPPVKLPLPLPAPGKSTPPPEEEKPTPPTPVKSSPPPEKSL 689
Qy   260  PTPTTPKEPAPTTKEPAPTAPKKPAPTTPKEPAPTTTKEPSPTT                319

Db   690  PPPTLTTSPPPQEKPTPPSTPSKPPPPPSPVETLPPPESKSSPPPEEVSSPPPQOAPKSSSP- 747
Qy   320  PKEPAPTTIKSAPTITKEPAPTTTKSAPTTTKEPSPTTTKEPAPTTPKKPAP        379

Db   748  PAPVS---SPPPLKSSPPPVPESSPPPTTKSSSPLAPVSSPPPTTKEPAPTTTPKEPAPTTP 802
Qy   380  TTPKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTTPKKLTPTTPEKLAPTTPE       439

Db   803  PTPKSSPPLAPVSSPPQVEKTSPPPAPVSSPPPTPKSSPPLAPVSSPPPTPKEPAPTTPKE- 862
Qy   440  KPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKSSPPPEVSSPPPV  499

Db   863  SSPPLEPKPSSPPPSVSSPPPTTVKSSPPPTMTPKSSSPPPAHVSSPPPEAEKSSPP      922
Qy   500  TTPKETAPTTPKGTA---------PTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSD   548

Db   923  LAPISSPPSEPKSPSSPPPMVEKTSPPPATVSSPPPAIS--------PKETAPTTPK---   979
Qy   549  KPAPT--------TPKGTAPTTP--KERAPTTPKEPAPTTTK-GTAPTTLKEPAPT      593

Db   980  PPAPVSSPPPTPKPLPPPAPVSSPPPPVVKSSPPPTPKPLPPPPPVSSPPPT          1039
Qy   594  TPKKPAPKELAPTTTKGPTSTISDKPAPTTT---PKETAPTTPK---EPAPTTPKKPAPTT 647

Db  1040  XKPLPPP---APVSSPPPVVKSSPPPAPVSLPPPTPTRKSSPPPRIRVSSPRPVVKCCPPTL 1096
Qy  1040  XKPLPPP---APVSSPPPVVKSSPPPAPVSLPPPTPTRKSSPPPRIRVSSPRPVVKCCPPTL 1096

Db   648  PETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA   707
Qy  1097  VSSPPPAPKSLPPPTPVSSP----PPEVKSSPPPTPVSSPPPAPKSSPPPTPVS       1146

Db   708  TKPEMTTTAKDKTTERDLKTTP----ETTTAAPKMTKETATTTEKTTESKITATTTQVT   762
Qy  1147  SPPP--------ELKSSPPPAPVSSPPSAPKSSPPPEVKSSPPPAPIS             1194

Db   763  STTTQDTTPFKITILKTTILAPKYTTYTTKTITTTEIMNKPEEFAKPKDRATNSKATTPKP 822
Qy  1195  SPPPPAKSP--------PMSSLPPPVKSSPPPAPVSSP                       1228

Db   823  QKPTKAPKKPTSTKKPKTMPVRVRKPKTTPTPRKMTSTMPELNPTSRIAPAMLQTTTRPNQ 882
Qy  1229  PPPMKSSPPPAPISSPPPAP--VKPSLPPPAPVSSPPPAV                     1271

Db   883  TPNSKLVEVNPKSEDAGGAGEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD  942
Qy  1272  -----------PKKEEDSTA                                         1302

Db   943  E  943
Db  1303  K 1303
```

ID   Q20007   PRELIMINARY;   PRT;   1274 AA.
AC   Q20007;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   COSMID F35A5.
DE   F35A5.1.
GN   F35A5.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Leimbach D.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U46675; AAB52641.1; -.
SQ   SEQUENCE 1274 AA; 138065 MW; 07F6B0D292C1799F CRC64;

Query Match   14.8%;   Score 802; DB 5; Length 1274;
Best Local Similarity   28.0%; Pred. No. 2e-44;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

```
Qy     6  KNRTKK-KP-----------TP------KPPVVDEAGSGLDNGD----FKVTTPDTSTTQHNKVSTS  50
Db   274  KNPTKKWKPPWEDDETPVEEVKEPPVPEKKAPVLKKKDPAPAAKARDPSPSKAAPKKVEPS        333

Qy    51  PKITTAKPI-NPRPSLPPNSDISKETSLTVNKETTVETKETTTNKQTTSDGKEKTTSAK        109
Db   334  SPVVPPTPVKNPVKKYKPPWEVDDEPAEEVKRPSAPEKKTPVLKRKEPEPSSTTPSSDPS       393

Qy   110  ETQSIEKTSAKDLAPTSKVLAKPTPKAE--------------LTTP               148
Db   394  PKKAAPAVKPRDSSPKKATPLQADPKAQGEVPPTPVKNPVKKYKPPWEVDDEDEVEEVKQP       453

Qy   149  KEPTPTTP-----KEPASTTPKEP---TP-TTIKSAPTTPKEPAPTTTKSAPTTPKEPAP  199
Db   454  EAPAKKTPVLKRKEPAAKDTAKPATSKTPETPEKKDPVKPRDSSPKKVAAKPDSAQAPA-      512

Qy   200  TTTKEPA-----------------PTTPKEPAPTTTKEPAPT-------TTKSAPTTP    233
Db   513  TPVKNPEVKKWRPPWEDDETPADDVSKPTDAKKTPSLAKKDPAPAKAKESLKPKADTKAAPAKP 572

Qy   234  KEP-----APTTPKKPAPTTPKE-----PAPTTPKEPTPTTPKEPAPTTTKEPAPTTPK-EP 283
Db   513  TPVKNPEVKKWRPPWEDDETPADDVSKPTDAKKTPSLAKKDPAPAKAKESLKPKADTKAAPAKP 572
```

RESULT 11

```
Query Match             14.8%;  Score 800.5;  DB 10;  Length 1480;
Best Local Similarity   29.6%;  Pred. No. 2.9e-44;
Matches  307;  Conservative  54;  Mismatches  505;  Indels  171;  Gaps  45;
```

```
Db   573 RDPSPKKVAPTAPEKKTPVLAKKEPAGPADSKTKEPEKSKPRDPSPKKAVPAKPVPKTEV 632
Qy   284 APTPAPKKPAP---------------------SPTTPKEPAPTTTKSAPT 333
Db   633 APAAAVKKPEPISKPKDTAPKKAEPNSPVVP-PTPVKNPVKKWKKPPWEDDDAPAKPVSLPE 691
Qy   334 TTPKEPAPTTTKSAPTTP--KEPSPTTTKEPAPTTPK--EPAPTTPKKPAP--TTPKEPAPT 388
Db   692 PEKK-TPVLAKKAPTKPDSEAAADPVSGPSSKDPKLAKKAPVKPRDPSPMKAVPIKPAPK 750
Qy   389 TPKEPAPTTTKKPAPTTA-PKEPAPTTPKETAPTTTPKKLTPTTP------EKLAPTTPEKPA 442
Db   751 T--EVPAVVKKPEPVAKSRDPSPKKAK-AEPNSP--VVPPTPVKNPVKKWKKPPWEDDDA 805
Qy   443 PTTPEELAPTTPEEP----TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTT-PKE 496
Db   806 PAEPVNVPEPEKKTPVLAKKTPVKRDPSPKKAVPAKPSTKTDAPPVSVKKPEPVSKPKE 865
Qy   497 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT-TTKEPTSTTSDKPAPTTP 555
Db   866 PSPKKAEPNSPVVP----PTPVKNPVKKW-KPPWEDDDEPTEEVKKPSE--PEKKTPVLA 918
Qy   556 KGTAPTTPKEPAPTTPKGTAPTTLKEPAPTT----------------KEL 603
Db   919 K-KEPEKPKD-APKVAAKPRDPSPKKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEA 974
Qy   604 APTTKGPTS---------TTSDKP----------APTTPKET------APTTPKEPAPTTPKKPAPT 646
Db   975 PPTPVKNPVKKWKKPPWEDDDEPAEPVSAPEPEKKITPVLAKKAPAKPRDP--SPKKAAPV 1031
Qy   647 TPETPPPTTSEVSTPTTKEPTTIHKSP----DESTPELSA-EP---TPKALENSPKEPG 698
Db  1032 AAK-PDPKIPEV-PPTPVKNPVKKWKKPPWEDDDEPSEPVSAPEPEKKTPVLAKKAPTKPA 1089
Qy   699 V----------PTTKTPAAT------KPEMTTTAKDKTTERDLRTTPETT--TAAPK 737
Db  1090 TKPDSEAAADPVSGPTSKDPKLSKKAPVEKPKPTTDPKDDKLKEPSPAKKPEKAPEPAAPK 1149
Qy   738 MTKETATTTEKTTESKITATTTQVTSTTQODTTPFKITTLKTTTLAPKYTTTKTITTTE 797
Db  1150 KWKPVWDDDDPDPEPEADFTVPAPSKKPDTEDPADPLG-----GPKTKDPK------ 1193
Qy   798 IMNKPEETAKPKDRATNSKATTPKPQKPTKAPKPTSTKKPKTMPRVKRPK-------- 848
Db  1194 -LNKKAPAEKPTEK--------PKPKEVSKEPRAPTERPKP-AAPKKKWKPPWEDDPDEPE 1243
Qy   849 ---TTPTPRKMTSTMP 861
Db  1244 ADFTMPAPKKPDTEDP 1259
```

```
Qy    10 KKKPTPKPPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTS----PKITTAKPINPRPSL 65
Db   117 KPRPHPKPPNVK------------PHPHPKPPTKPHPHPKPPTKHHPHPKPPKPPSV 167
Qy    66 PPNSDTDSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTTSAKETQSIEKTSAKDLAPT 125
Db   168 KPPPSTPK-----------------------PPTTNPPPSTPQPPFHKPPPCTTPTTPPVASPPMATPP 211
Qy   126 SKV--LAKPTPKAETTTKGPALTTPKEPTP----TTP--KEPASTTPKEPTPTTTIKSAPTTP 179
Db   212 TQMPIATP-PIAKSPVATPPIATPPTAAPPIITILPPPVAVPPIITTPPIANPPIMPPIATP 270
Qy   180 KEPAPTTT-----KSAPTTPKEPAPTTT---------------KEPAPTTPKEPAPTTKE 220
Db   271 PVAAPPIITNPPPISKPPVTTP----PTTTPPIIAKPPIAIPPISTPPAATPPAAITTLP 326
Qy   221 PA--------PTTTKSAPTTP-KEPAPTTPKEP----SPTTTKEPAPTTPKEPAP 270
Db   327 PAKPPVAISPIVI--PPVTPIAQPPVATPPIATPPVAIPPIAIPPTSKSPISPISESP 384
Qy   271 TTTKEPAPTTP-KEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTT---KEPSPTTPKEPAP 325
Db   385 VATPPTATSPIKTPPAKPPVATPPIAKSPIATPPATPPIEKPPVATPPVATPPP 444
Qy   326 TTTTKSAPTTTKEPAPTTTKSAPTTPKEP--SPTTTKEPAPTTPKEPAP 380
Db   445 TAT----PPVAKPVETPPIATPPAKPPISTPPISKPPVATPPAATPPITTPPVKPPVA 501
Qy   381 TPKEPAPTTPKEP---APTTTKKPAPTAP--KEPAPTTPKETAPTT----PKKLTP--T 428
Db   502 TPPLAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPIAKPPVVTPPTT 561
Qy   429 TPEKLAPTTPEKPAPTTPEELAPTT-PEEPTPTTPEEPAPTTPKAAAPNT---PKEPAPT 484
Db   562 TPPIATPPVAKPVATPPIATPPTAKPPISTPPISKSPVATPPAATPPIITPPPAKPPVA 621
Qy   485 TPKEPAPTTPKEPAPTTPKETAP---TTPKGTAPTT---LKEPAPTTPKKPAPKELAPTTT 539
Db   622 TPPIATPPIAKPPVATPPIATSPVAKPPVATPPIKTPPAKPPVAIPPIATPPVA 681
Qy   540 KEPTSTTSDKPAPTTPKGTAP--TTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKP 598
Db   682 KPPVAT----PPIATPPIATPPVVTPPIATPPVATPPIAKPPTI--PPTATPPVA 736
Qy   599 APKELAPTTTKGPTST--TSDKPAPTTPKETAPT--TPKEPAPTTPKKPAPTTP-----E 649
Db   737 MPPIATPPTAKPPIATPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVA 796
Qy   650 TPPPTTSEVSTPTTTKEPTTIH--KSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA 707
Db   797 TPPIAKPPVATPPTAPPVVAKPPVATPPIATPPTAKPPIILTPPISKSPVATPPAA 856
Qy   708 TKPEMT-TTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTT 766
Db   857 TPPIITPPPAKPPVATPPIATPP--IAKPPVATPPIATPPIATSPVAKPPVAIPPIKTPP 914
Qy   767 QDTTPFKITTLKTTTLAPKYTTTKKT----ITTTEIMNKPEETAKPKDRATNSKATTPKP 822
```

```
Db   915  PAKPPVAIPPIATPPVAKPPVATPPTAPPIATSPIATPPVVT---PPTATSPVATPPIA  971
Qy   823  OKPTKAPKKPISTKKPKPKTMPRVRKPKTTPTPRKMISTMPELNP----------TSRIAE---  871
Db   972  KPPTTTP--PTAT-PPVAMPPIATP----PTAKPPVAIPPIANPPVEKPPVATPPIAKPPT  1025
Qy   872  -----AMLQTTTRPNQTPNSKL---VEVNPKSEDAGGABGEIPHMLLRPHVPMPEVTPDMD  924
Db  1026  VLPPIAKPPVETSPTATPPTATPPVAIPP--------------VVKPPVAIPPIIKPPV  1070
Qy   925  YLPRVPNOGIIINPMLS  941
Db  1071  AtPPVTNPPTAMPPIVT  1087
```

RESULT 12
Q41805
```
ID   Q41805;     PRELIMINARY;     PRT;   1188 AA.
AC   Q41805;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z34465; CAA84230.1; -.
DR   Mendel; 14346; Zeama; 2368;14346.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003591; LRR.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   SMART; SM00370; LRR; 4.
KW   Signal.
FT   SIGNAL       1     27       POTENTIAL.
SQ   SEQUENCE   1188 AA;  120980 MW;  2C77C7F8D7130149 CRC64;
```

```
Query Match         14.5%;    Score 784;    DB 10;    Length 1188;
Best Local Similarity  29.9%;   Pred. No. 2.8e-43;
Matches 230;    Conservative 59;    Mismatches 359;    Indels 122;    Gaps 33;
```

```
Qy   13  PTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVYSTSPKITTAKPINPRPSL----PPN   68
Db  460  PTPHSPPAD--------DYVPPTPVPGKSPPATSPSPQVQPPAAStPPPSLVKLSPPQ  510
Qy   69  SDTSKETSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV  128
Db  511  APVG-------------SPPPPVVKTTSPAPIG---SPSPPPVSVV  541
Qy  129  -----LAKPTPKAETTTKGPALTTPKEPTPTTPKEPTPTTIKSAPTTPKEPA  183
Db  542  SPPPPVKSPPPPAPVGSPPPPPAVSPPPPPAVSPPPPVKSP--PPPPLVASPPPPVKSP  599
Qy  184  PTTTKSAPTtP---KEPAPTTKEPAPTTPKEPAPTTKEPAPTTP-KEPAPTT  240
Db  600  PPAPVASPPPPVKSPPPPTPVASPPPPMKSPPPPTPVSSPPPPEKSPPPPP  659
Qy  241  PKKPAPTTPKEPAPTT------PKE---PTPTTPKEPAPTTKEPAPTAPKKP  291
Db  660  PAKSTPPPEEYPtPPTSVKSSPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP-PSSPEKP  718
Qy  292  APTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTKSAPTTTKEPAPTTIKSAPTTPK  351
Db  719  SP--PKEPVSSPPQTP----KSSPPPAPVSSPPPTPVSSPPALAPVSSPPVSKSSSP----  768
Qy  352  EPSPTTTKKEPAPTTPKEPAPTTPKEPAPTTKEPAPTTPKKPAPTAPKEPAP  411
```

```
Db  769  PPAPLSSPPAPQVKSSPPPVQVSSP-PPAPKSSPPPLAPVSSPPQVEKTSPPPAPLSSPP  827
Qy  412  TTPKETAP-------TTP----KKLTPTTPEKLAPTTPEKLAPTT  458
Db  828  LAPKSSPPHVVVSSPPPPVVKSSPPPAPVSSPPPLLTPKPASPPAHVSSPPPEVKPSTP--PA  885
Qy  459  PTTPEEPAPTTPKAAAPNTPKE-PAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTT  517
Db  886  PTTVISP-PSEPKSSPPPVSLPPPIVKSSPPPAAVSSP-PMTPKSSPPPVVVSSPPPT  943
Qy  518  LKE---PA----PTTPKK---PAPKELAPTTKE---PTSTTSDKPAPTTPKGTAPTTP  563
Db  944  VKSSPPPAPVSSPPPAPTPKSSPPPAPVNLPPPEVKSSPPPPTPVSSPPPA--PKSSPPPAP  1000
Qy  564  -KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTLKKPAPKELAPTTTKGPTSTTSDKPAPT  622
Db 1001  MSSPPPPPVEVKSPPPPAPVSSPPPPVKSPPPPAPVSSPP--PPVKSPPPPAPVSSPPPPV  1057
Qy  623  -TPKETAPTTPKEPAPTTTPKKFKPAPTTPEETPTPPTTSEVSTPTTTKEPTTIHKSPDESTPEL  681
Db 1058  KSPPPPAPISSPPPPVKSPPPPAPVS--SPPPEVKSPPPPAPVSSPPPPIKSPPPPAPVS  1115
Qy  682  SAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPET  731
Db 1116  SPPPAP------VKPPSLP-PPAPVSSPPPPVVTPAAPPKKEEQSLPPPAES  1158
```

RESULT 13
P70670
```
ID   P70670;     PRELIMINARY;     PRT;   2187 AA.
AC   P70670;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE   MUSCLE-SPECIFIC FORM GP220).
GN   NACA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96312450; PubMed=8698236;
RA   Yotov W.V., St-Arnaud R.;
RT   "Differential splicing-in of a proline-rich exon converts alphaNAC
RT   into a muscle-specific transcription factor.";
RL   Genes Dev. 10:1763-1772(1996).
DR   EMBL; U48364; AAB18734.1; -.
DR   EMBL; U48363; AAB18732.1; -.
DR   MGD; MGI:1060095; Naca.
DR   InterPro; IPR002715; NAC.
DR   InterPro; IPR003037; TS-N.
DR   Pfam; PF01849; NAC; 1.
DR   Pfam; PF02094; TS-N; 1.
SQ   SEQUENCE   2187 AA;  220599 MW;  003646AA864DEBFD CRC64;
```

```
Query Match         14.3%;    Score 776;    DB 11;    Length 2187;
Best Local Similarity  27.7%;   Pred. No. 1.6e-42;
Matches 288;    Conservative 122;    Mismatches 389;    Indels 240;    Gaps 49;
```

```
Qy   34  VTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKE------  84
Db  806  VQSPKVDPIMSDVTPISPKKTSATAV-PK------DTSATLSLKSVPAVTSLSPEKAPV  857
Qy   85  VETKEKTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPA  144
Db  858  APSNEAITVPTIELPTSLKNALAAATPKETLATSIPKVTSPS---PQKTPKSVSLKGAPA  913
Qy  145  LTT------------PKE------PTPTTKEPASTT----------PKEPTPTT  171
```

```
Db   914 MTSKKATEIAASKDVSPSQPKEVPLLQHVPPTSPPKSPVSDTLSGALTSPPPKGP-PAT 972
Qy   172 IKSAPTTPKE-PAPTTTKSAPTTPK------------------EPAPTTTKEPAPTTPKEPAPTTKEP 221
         | | :  :|:|:|             |||||  :| :: :|:| ||
Db   973 LAEPTIYPKKSPKPAASKKTPATPSPEGVTAVPLEIPPCSKKAPKIAAPKESSATSSSKR 1032
Qy   222 APTT--TKSAPT----------------TPKEPAPTTPKK--PAPTTTPKEPAPTTP 257
         |||    |: :                  |  |||:| |||
Db  1033 APKTAVSKEIPSKGVTAVPLEISLPLKETSKSAATPGEKSASSPKRSPKTAGPKE--- 1089
Qy   258 KEPTTPTTPKEPAPTTKEPAPTTPKEPAPTTARKKPAAPTT-----PKEPAP----TTPKEPAPTT 311
         :|:|:||:|||||||||||:| ||:|:|                  |||||||    |||||||||
Db  1090 GGVTAVPPEISLPPKETPQNATPNESLAASSQKRSPKTSVPKETPPGGVTAMPLEIPSAP 1149
Qy   312 TKEPSPTTPK-----EPAPTTTKSAPTTTKEPAPTTT--KSAPTTPK---------EPSPTT 357
         |||| ||||      ||||||| | ||| ||||||   |||| |||         ||:|||
Db  1150 QKAPKTAVPKQIPTPEDAVTILAGSPSLSPKKASKITAAPKEPAPTPSVGVTAVSGEISPSP 1209
Qy   358 TKEPAPTTPKEPAPTTTPKKPKPAPTT---PKE-----------PAPTTPKEPAP------T 396
         |||||||||||||  :|:|| || :||    |||           |||||||||||      |
Db  1210 KKTSKTAAPKENSATLPPKRSPKTAAPKEPATSSEGVTAVPSEISPSPPPPASKGVPVT 1269
Qy   397 TTKKKPAPTTAPKEPAPTTPKETAPTTTPKKLTPTTPKEKLAPTTPKK--PAPTTPEELAPTTPE 455
         ::|| || |::|||||||| |:| |||||| :|| ||||  |||||||  ||||| ||:|||||
Db  1270 LTPKGAPNALAE-SPASPKKVPKIPAAPEETSTTP----------SPQKIPKVAGPKEASATPS 1322
Qy   456 EPTPTT--------------PEEPAPTTPKAAAAPNTPKEPAPTTPKE-PAPTTTKKEPAPTTPKETA 506
         ||| ||               ||| |||| | : | |:| ||||||| |||||:| |||||||:|
Db  1323 KKTPKTAVPKEKETSAPSEGVTAVPLEIPPSPRKAPKTAAPKEKETPAPS---PEGATTAPVQIP 1380
Qy   507 PTTPKGTAPTTLKEPAPTTP------------------KKPAPKELAPTTTKEPTS-- 544
         ||||||| ||| ||||||||                  |:|||| || |||||:|
Db  1381 PSPRKGSKAGSKE-TPTTPSPEGVTAAPLEIPSSKKTSKSMASPKEILVTPSSKKLSQT 1439
Qy   545 ----VGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-A 1496
         ...
Db  1440 VGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-A 1496
Qy   600 PKELAPTTTKGPTSTSDKAPTTTKEPAPTTPKEPAPTTPKKPAPTTPETP-----PPT 654
         |||||||||||||  | |||| |||||||||||||||||| |||||| |      |||
Db  1497 PKTAAPPSER--VTTVPPEKPA-TPQKASGTTASKVPVPAEETQEVAVSSRETPVTPAVPPV 1554
Qy   655 TSEVSTPTTTK------EPTTIHKSPDESTPELSAEPTPKALENSPKE--PGVPTTKTPAA 707
         :| ||||||       ||| | |   |  ||:|||| :|||||| |   ||||| ||||
Db  1555 KNPSSHKKTSKTLELKEAPATLPPSPTRKSPKIPSSKKAPRT--SAPKEPPASPSIK-PVT 1611
Qy   708 TKPEMTTTAKDKTTERDLRTTPETTTNAAPKMTKETATTTEKSKITATTTQVTSTTTQ 767
         |||| |:| || :||  | ||||  | | |  |:|   ||:|| :|:| || |:|||
Db  1612 T--SLAQTAPPSLQKAPSTTIPKENLAAPAV--LPVSSKSPAAPARASASLSPAAPAAP 1665
Qy   768 DTTPFKITTLKTTLAPKVTTTRKTITTTEIMNKPETTAPKPKDRAINSKATTPKPQKPTK 827
         |  | |: |  || ||:|  ||:: ::: :  ||||  | |    |  ||||| |||||
Db  1666 QTAPKEATTIPSCKKAAATETFIETSTAPSLEGAPKETSE---TSVSKVLMSSP---- 1716
Qy   828 APKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSK 887
         ||  |:|||  |:  | |: | || | |::|  | :    |||:|    |  |   ||:|
Db  1717 -PKKASSSKRASTLP--------------------AITLPSLKEASVLS---------- 1752
Qy   888 LVEVNPKSEDAGGAEGETP 906
         ::: |:||    ||  |||
Db  1753 DSHISPVS-DACSTGTTTP 1770
```

```
Query Match        13.6%;  Score 739;  DB 5;  Length 2112;
Best Local Similarity   27.2%;
Matches  281;  Conservative  98;  Mismatches  339;  Indels  316;  Gaps  44;

Qy     4 NKKNRIKKKKPPPKPPVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRP 63
         |   :  |:|             |  |||: | || ::    |    ||| ||  || |
Db  1023 NHPDQIQTKPLKCKKVVSQGGSSSN------STSNSSSSSNNSGGSSSSNSGGSGSSSS---- 1070

Qy    64 SLPPNSDTSKETSLTIVNKETTV--ETKETTTTNKQTS---TDGKEKTTTSAKETQSI-- 114
         | ||::  :  ::|   ||:  |    |   | |||    |    |    |  |||
Db  1071 SSSSNSGSSSNTGSSSNSGASSSGGSSNSGSSSNSGSSNNSGGSSNSGCNQSTTSSSSTSSSSSSSS 1130

Qy   115 ----EKTSAKDLAPTSKVLAKPTTPKAETTTKGP----------- 143
         :   |:|:  | ||| | |:|||:|
Db  1131 NNNNQGSSSSSSSSSSSSSSSSSSSSSKPNPSETCKVNGQFIGDRSDCAKFYRCVDNDRGGFNMVPF 1190

Qy   144 --------------ALTTPKEPPTTPKEPASTTPKEPPTTIKSAPTT 178
```

```
Db  1191  SCGPGTVWDAQMQACNHAWAVKECGGIAPPTTSTPTTSR-PTTASTSRPSDQTSTSRPTG  1249
               ||||  | | :           |  |                         |  |||
Qy   179  PKEPAPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTT----------------------   217
Db  1250  P----PTTARPVTARPTSSPTTASSSQTTSPVTQQAPNTDGKCRSEGFMADPNNCSKFYR  1305
               |||                                                       
Qy   218  ------------TKEP----APT-------------------------------------   218
Db  1306  CVRNNKGGFTSIPFQCGAGTVWDQDLQTCNHINFNNCSTGTTESTTPKPPC--EPATNGTTA  1363
                                                  :  |||  : :|   |  ||| ||
Qy   249  PKEPAPTTPKEPTTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTAPKKPAPTTPKEPA   308
Db  1364  TSTSSTTTP--PPTTTDLPPISTTGLP-PTTTTELPPT------------TTTDLPPTTTRLPP  1413
          : | |   |      ||||:|      |  |||   |            | |  |||| | |:|
Qy   309  PTTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK---EPSPTTTKEPAP--   363
Db  1414  TTTTTSLPPTTTTGLPPTTTTLSSETETSTVTISPESTTQQPPSTTTMKPLPAG        1473
          : ||||     |:      |||  : |  ||  : :|||     ||||  |  |
Qy   364  ----------------------------------------TTTPK-----EPAP---TT   373
Db  1474  TECTGEGYMADPEDCRKYYRCINAGASYRKYNFTCPKGTGWNEEVQTCDYVENIPRCSKL  1533
                                                    |||
Qy   364  ----------------------------------------TTTPK-----EPAP---TT   373
Db  1534  PAEPITTTPSEESKDPGSSTPQSTDEPTTVTKPI-TKPTE-EPSTEKPQKPTTQYPEKPT  1591
              :  | ||||||  :    | | |  : | :|   ::|   |  |  |  ||   |||
Qy   429  TPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTPKEPAPTTPKEAAAPNTPKEPAPTTPKE   488
Db  1592  TTEE--PEKPQKPT-TTEYPQKPTTEEPTPEKPQKPT-TTEYPQKPTTTEEPTPEKPQK   1647
          : :|    | :| |        |  ||    | :| |        |   ||    | :|
Qy   489  PAPTT--PKEPAPTTPKETA------PTTPK--GTAPTTLKEPAPTTPKRPAPKELAPTTK   540
Db  1648  PTTTEYPQKPTTTEEPTTTSIPGYVNPTTISVPGINPTTTIPIPVETTISTPGYK---PTTTG  1704
                :|  |||  | ||:   ||  |: : |  |  ||     || |  :|  |      |||:
Qy   541  EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKK-   597
Db  1705  EPITTTTT-LPSTTTDAIQEPTTSKKPEPTTTTTESPESSTPEGSVTTLQPEPQPNYNCSSE  1763
              ||||     |   |   ||  | |  |          |  |  |   | :|  | | |
Qy   598  ---PAPKELA----------PTTTKGPISTTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTP   648
Db  1764  GFPPDPEDCCSRYYRCVDAAKNGKYQVYAFRCGKGTGVWDISTETCNYADQVSGN-----  1816
                                  |                  :
Qy   649  ETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAT   708
Db  1817  --CSSGQTTTPGTTTPEPGTGTTESTTSSGKP-------ETTSKAPEN------TTTWA--  1857
                :  ||     |  |  | ||          |            | |
Qy   709  KPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQD   768
Db  1858  -PETTTT---------SSPETTTT---VASETTTTSGT-------TTTATPETTTKP      1895
          | |||              ||||        |||         ||  |  ||||
Qy   769  TTPFKITTTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKA   828
Db  1896  PKP-----------ETTTTIAGEETSTSKSPTTTE------------SPAPSTNTSA      1928
          |               ||  |  |  ||  |||                |   |
Qy   829  PKKPTSTKKPKTMP  842
Db  1929  P-----CPETGP  1935
```

```
Query Match            13.2%;   Score 715;   DB 2;   Length 763;
Best Local Similarity  28.2%;   Pred. No. 5.7e-39;
Matches 248;  Conservative 39;  Mismatches 361;  Indels 232;  Gaps 39;

Qy   132  PTPKAETTTKGPALTTPKEPTTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP   191
               | |  | |  |   | | | ||  ||   ||  |  | |       | |||
Db     3  PVP------APRALAPLPPAPAPAEPKSKPPFFPAP----------------PAPPCWMLVSAAP    46
Qy   192  TTPKEPAPTTTKKEPAPTTPKEPAPTTPKEPAP-----TTTKKPAPTTAPKEPAPTTPKKPAPTTPKE   251
              | |  | |     ||  | | |          |                  | |  |
Db    47  PCP---PAP------PAPPKSKAPFPPVPAPPAREIAPPLP--PAP-------PEAPRE    90
Qy   252  PAPTTPKEPTPTT-----PKEPAPTTPKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK   305
              | |  | |      | |   |||  | | |    |  | |       | |  | |
Db    91  SRPALPPCPPPPVVIPDPEPEAPAPVPPAPNSPFPFPPPAPKFVPRAPPVP--PVPNSPP   148
Qy   306  EP--APTTTPKEPSPTTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAP   363
              :|             |  | |        |     |    |   | |     | |
Db   149  FPPPPPAALNPPAP-----PAPPLANSPPLPPAPPPAGT--PPAAPWPPVPAAAPKSKPA   201
Qy   364  TTPKEPAPTTPKRPAPTTPKEPAPTTPKEPAP-----TTTKKPAPTTAPKEPAPTTPKETAP   419
              | |  | |   | ||  | | |              |    |       | |  |
Db   202  PMPATPMEFPPLPPVPPPDPISKETPPAPPAPPIPPAPVPPPVPPLP   256
Qy   420  TTTPKKLTPTTTPEKLAP-----------TTTPEKPAPTTPEELAPTTPEEPTTTPEEPAPTTT   469
              : : |             :                  | |   | |
Db   257  PVPNIPPPAPP---APPVAVAVAVLVAPCPPLPPLPNNHPPAPPAAPVPGVPLAPLPNSHP   313
Qy   470  ---PKAAAPNTPKEPAP-------------------------------------------   483
               |  |  | |||
Db   314  PAPPSAPVPGVPLAPLPISGRPVSVWKGSFTTLSTFCRVCSGEVLAGALNPSRSPSPL   373
Qy   484  --TTPKEPAPTTPKEPAPTTPKEPAPTTPKETA----PTTPKGTA---PTTLKEPAPTTP-KKPAPKE   533
              | |  | |  | |  | |           | |        |        | |  |
Db   374  TTTTTALPAPIPPLPPLPLPINTAVPPIPPLPPVTALAPPLPPLAPPLPISPGVPPAP--   431
Qy   534  LAPTTTKEPTSTTTSDKPAPTTPKGTPAPTTTPKGTAPTTPKEPA---PTTPKEPAPTTPKGTAPTTL   587
              | |   | |    |   | |       | |  | |       | |  | |
Db   432  --PIPPGKPWTTPPLPAPAPEPK-TVPVLPPGSCPPSEKPNPAPPEPEPEPKSSPALPP   488
Qy   588  KEPAPTTP---KKPAPKELAPTTTKGP-TSTTSDKPAPTTPKET---APTTPKEPAPTTP   640
              | |       | |  | |  |          |  | |        | |  | |
Db   489  APPAPSMPSAVRVPPSPPIPAPAAPRASMPALPPAPSPPAATRLCPPLPPSSPPAPNSP   548
Qy   641  KKPAPTTPETPPPTTISEVSTPTTTKEPTTIHKSP----DESTPELSAEPTPKA--LENSP   694
              | |   | |               | |            |  |  | |
Db   549  --PAPPAPTPPPKLLS--ANPPCPPVPPAPNRPPAPPAPPELPAPPDDPPTPPVANSP   604
Qy   695  KEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKI   754
               |   | |   |  |         |  |
Db   605  --PAPPAPPAPPSALPFVNPPA--------------------PPTPAAPK------SRPAL   637
Qy   755  TATTTQVTSTTTQDTTTPFKITTTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATN   814
               |   |  |  |     | |||
Db   638  PAAPPAPPAPPVRATTP-----------------------PPAPAPPAPN   665
```

```
Qy  815  SKATTPKPQKP-----TKAPKKPTSTKKPKTMPRVRKPKTTPTTPRKMTSTM------PELNP  865
             | | | | |          |  |  | |:| |        | | | |              |:    | | |
Db  666  SMALPPAPPDPPIPLLATPPAPPLPMSPPAPPLPPAAPDPPAPPLTINQPPSPPLAP  725

Qy  866  TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET  905
             | |         |:        | : |:|
Db  726  VPGAPLAPLPINGRPVFARKNSLI---------GSSSGDT  756
```

Search completed: April 26, 2002, 16:22:47
Job time: 552 sec

```
                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:          April 26, 2002, 16:22:02 ; Search time 153.5 Seconds
                                             (without alignments)
                                             922.421 Million cell updates/sec

Title:           US-09-556-246-1_copy_200_1167
Perfect score:   5155
Sequence:        1 VKDNKKNRTKKKPTPKPPVV.........GKPVDGLTTLRNGTLVAFRG 968

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters:   473505

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :       SPTREMBL_17:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10: sp_plant:*
                 11: sp_rodent:*
                 12: sp_virus:*
                 13: sp_vertebrate:*
                 14: sp_unclassified:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5155	100.0	1404	4	Q92954	homo sapien
2	5146	99.8	1404	4	Q9BX49	homo sapien
3	2046	39.1	1054	11	Q9MM99	zea mays (m
4	967	18.8	1079	5	Q9N457	caenorhabdi
5	936.5	18.2	1049	5	Q91TS1	drosophila
6	865.5	16.8	1795	5	O76894	drosophila
7	853	16.5	1489	10	O96449	phytophthor
8	808.5	15.7	1315	10	Q9SPM0	zea mays (m
9	802	15.5	1274	10	Q20007	caenorhabdi
10	800.5	15.5	1480	10	Q9LIE8	arabidopsis
11	784	15.2	1188	10	Q41805	zea mays (m
12	776	15.1	2187	11	P70670	mus musculu
13	739	14.3	2112	11	Q9VEL9	drosophila
14	715	13.9	763	2	Q9XDH2	drosophila
15	707.5	13.7	555	10	Q9PPQ6	chlamydomon
16	671	13.0	1151	13	O57580	gallus gall
17	667.5	12.9	4880	11	Q9JLT1	rattus norv
18	667.5	12.9	5085	11	Q9JKS6	rattus norv
19	660.5	12.8	6677	5	Q9N435	caenorhabdi
20	659.5	12.8	1229	5	Q94185	caenorhabdi
21	657.5	12.8	2284	5	Q9VPG1	drosophila
22	653.5	12.7	1514	5	Q9GUW7	leishmania
23	647	12.6	4833	11	Q9QYX6	mus musculu
24	647	12.6	5038	11	Q9QYX7	mus musculu
25	641.5	12.4	2089	4	Q90YX7	homo sapien
26	633	12.3	489	10	Q14676	homo sapien
27	632.5	12.3	489	4	Q41707	vigna ungui
28	632	12.3	761	10	Q10465	homo sapien
29	629	12.3	6632	5	Q92Q10	arabidopsis
30	626.5	12.2	990	13	Q17362	caenorhabdi
31	626	12.2	3507	5	O23587	caenorhabdi
32	625	12.2	6642	5	Q91803	xenopus lae
33	624.5	12.1	1612	5	O01761	caenorhabdi
34	624.5	12.1	2768	5	O01761	caenorhabdi
35	622.5	12.1	839	2	Q9VQO2	drosophila
36	620.5	12.0	971	5	Q9VC00	drosophila
37	617	12.0	3570	4	Q9RX57	deinococcus
38	609.5	11.8	2344	5	Q9XVS4	caenorhabdi
39	607.5	11.8	801	5	O99552	homo sapien
40	607.5	11.8	924	12	Q9N3Y8	caenorhabdi
41	605	11.7	379	5	O23635	caenorhabdi
42	600.5	11.7	1893	5	Q99307	epstein-bar
43	600.5	11.6	409	10	Q9NKC9	drosophila
44	598.5	11.6	401	6	Q9SBM1	drosophila
45	584.5	11.3	956	10	Q9LJ64	arabidopsis

RESULT   1

ALIGNMENTS

```
ID   Q92954     PRELIMINARY;   PRT;   1404 AA.
AC   Q92954;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEGAKARYOCYTE STIMULATING FACTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.;
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA   Jacobs K., Turner K.;
RL   (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL   Mosher D.F. (eds.);
RL   Biology of vitronectins and their receptors., pp.45-52,
RL   Elsevier Science Publishers B.V. (1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Kriz M.J.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   Blood 78:279-279(1991).
DR   EMBL; U70136; AAB09089.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   InterPro; IPR002400; GF_cysknot.
DR   Pfam; PF00045; hemopexin; 2.
DR   Pfam; PF00086; Somatomedin_B; 2.
DR   Pfam; PF01033; hemopexin; 2.
DR   PRINTS; PR00438; GFCYSKNOT.
```

```
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
SQ   SEQUENCE  1404 AA; 151090 MW;  AABD7AD19B35F4F6 CRC64;

Query Match            100.0%;  Score 5155;  DB 4;  Length 1404;
Best Local Similarity  100.0%;  Pred. No. 0;  Mismatches 0;
Matches  968;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPIN 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPIN 259

Qy     61 PRPSLPPNSDTSKETSLIVNKETTVETKKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 PRPSLPPNSDTSKETSLIVNKETTVETKKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 319

Qy    121 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTTPKEPTPTTIKSAPTTPK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTTPKEPTPTTIKSAPTTPK 379

Qy    181 EPAPTTTKSAPTTPKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTI 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 EPAPTTTKSAPTTPKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTI 439

Qy    241 PKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    440 PKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 499

Qy    301 PTTPKEPAPTTTKEPSPTTPKEPAPTTPKEPAPTTTKSAPTTTKSAPTTPKEPSPTTTKE 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    500 PTTPKEPAPTTTKEPSPTTPKEPAPTTPKEPAPTTTKSAPTTTKSAPTTPKEPSPTTTKE 559

Qy    361 PAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTTPKPAPKEPAPTTPKETAPT 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 PAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTTPKETAPT 619

Qy    421 TPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 TPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 679

Qy    481 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 739

Qy    541 EPTSTTSDKEPAPTTPKGTAPTTPKKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 EPTSTTSDKPAPTTPKGTAPTTPKKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 799

Qy    601 KELAPTTPTTKGPTSTTSDKPAPTTPKETAPTTPEPPPTTSEVST 660
          ||||||||||||||||||||||||||||||||||||||||||||
Db    800 KELAPTTPTTKGPTSTTSDKPAPTTPKETAPTTPEPPPTTSEVST 859

Qy    661 PTTTKEPTTIHKSPDESTPELSAEPPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    860 PTTTKEPTTIHKSPDESTPELSAEPPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKT 919

Qy    721 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTQDTTPFKITTTLKTT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    920 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTQDTTPFKITTTLKTT 979

Qy    781 TLAPKVTTTKKIITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKT 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    980 TLAPKVTTTKKIITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKT 1039

Qy    841 MPRVRKPKTTTPPRKMTSTMPELNPTSRIAEAMLQTTTTRPNQTPNSKLVEVNPKSEDAGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1040 MPRVRKPKTTTPPRKMTSTMPELNPTSRIAEAMLQTTTTRPNQTPNSKLVEVNPKSEDAGG 1099

Qy    901 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1100 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN 1159
```

```
Qy    961 GTILVAFRG 968
          |||||||||
Db   1160 GTILVAFRG 1167
```

RESULT 2
Q9BX49
ID   Q9BX49;   PRELIMINARY;   PRT;   1404 AA.
AC   Q9BX49;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).
GN   BG174L6.2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133553; CAC36090.1; -.
SQ   SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;

Query Match   99.8%; Score 5146; DB 4; Length 1404;
Best Local Similarity 99.8%; Pred. No. 0; Mismatches 2;
Matches 966; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPIN 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPIN 259

Qy     61 PRPSLPPNSDTSKETSLIVNKETTVETKKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 PRPSLPPNSDTSKETSLIVNKETTVETKKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 319

Qy    121 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTTPKEPTPTTIKSAPTTPK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTTPKEPTPTTIKSAPTTPK 379

Qy    181 EPAPTTTKSAPTTPKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTI 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 EPAPTTTKSAPTTPKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTI 439

Qy    241 PKKPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTTPKEPA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    440 PKKPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTTPKEPA 499

Qy    301 PTTPKEPAPTTPKEPSPTTTKEPAPTTPKEPAPTTTKSAPTTTKSAPTTPKEPAPTT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    500 PTTPKEPAPTTPKEPSPTTTKEPAPTTPKEPAPTTTKSAPTTTKSAPTTPKEPAPTT 559

Qy    361 TPKKLTPTTPEKLAPTTPEELAPTTPEEPTPTTPEEPAPTTPKKPAAAPNTPKE 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 TPKKLTPTTPEKLAPTTPEELAPTTPEEPTPTTPEEPAPTTPKKPAAAPNTPKE 619

Qy    421 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 679

Qy    481 PAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 PAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 739

Qy    541 EPTSTTSDKPAPTTPKGTAPTTPKKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 EPTSTTCDKPAPTTPKGTAPTTPKKEPAPTTPKGTAPTTLKEPAPTTPKKPAP 799
```

```
Qy   601  KELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVST  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   800  KELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVST  859

Qy   661  PTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTPAKDKT  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   860  PTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTPAKDKT  919

Qy   721  TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   920  TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT  979

Qy   781  TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTTKPKPQKPTKAPKKPTSTKRKPKT  840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   980  TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTTKPKPQKPTKAPKKPTSTKRKPKT  1039

Qy   841  MPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1040  MPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG  1099

Qy   901  AEGETPHMLLRPHVEMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1100  AEGETPHMLLRPHVEMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRN  1159

Qy   961  GTLVAFRG  968
          ||||||||
Db  1160  GTLVAFRG  1167
```

RESULT 3

Q9JM99
ID   Q9JM99;   PRELIMINARY;   PRT;   1054 AA.
AC   Q9JM99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE   PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN   PRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ikegawa S., Nakamura Y.;
RT   "a novel mouse gene highly homologous to a human gene encoding
RT   megakaryocyte stimulating factor precursor and cartilage superficial
RT   zone protein.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034730; BAA92310.1; -.
DR   MGD; MGI:1891344; Prg4.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001212; Somatomedin_B.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ   SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match   39.7%; Score 2046; DB 11; Length 1054;
Best Local Similarity   45.9%;
Matches 449; Conservative 34; Mismatches 128; Indels 368; Gaps 20;

```
Qy     1  VKDNKKNRTKKKPPTPKPKPPVDEAGSGLDNGDFKVT--TPDTSTTQHNKVSTSPKITTAKP   58
          ||||||| |||||  |:|| |||||||||||||||||  |:|||||||||||:||||||||
Db   195  VKDNKKNTPKKKPNPEPPAVDEAGSGLDNGEFKLTPPPDPPDPPTPHSKVATSPKTTAAKP  254
```

Second column:

```
Qy    59  INPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTS  118
          : |||:|||||||||:|||| |:||| |||||||| ||||||||||||||||:|||||||
Db   255  VTPKPSLAPNSETSKEASLASNKETTVETKETTATNKQSSA-SKKKTTSVKETRSAEKTS  313

Qy   119  AKDLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTPTTPKEPASTTPKEPPPTTIKSAPTT  178
          |||:|||||||:|||||||||||||:||||||||||||| ||||||||||||||||:|||||
Db   314  DKDV---------------------------------------------------EPTSTTPK-  331

Qy   179  PKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKSAPTTPKEPAP  238
          |||| |||:||||||||||| ||||||  ||| |||||||||||||||||||||||||:
Db   332  TKKPV-TTTKESKFLP-----------------------LPQEPEPTTTAKEPEPTTTPKEPEPAP  376

Qy   239  PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKSAPTTPKEPSPTTT  298
          |||||||||||||||:|||||| ||||||||:||||||||||||||||||||||||||||
Db   377  TTPKEPEPTTTPKEPAPTTPKEPAPTTTPKEPAPTTPKKPAPTTAPKEPAPTTPKE      436

Qy   299  RAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT  358
          |:||||||||:||| |||| |||||||||||||||||||||||||||||||||||| ||
Db   437  PEPTTTKEPESTTTKREPEPTTTPKEP-----------------------EPTTKEPEPTTL  474

Qy   359  KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKPAPTAPKEPAPTTPKETA  418
          |||||||:||||||||||:|||||| |||||||||||||||||:||||||||||||||||
Db   475  KEPEPTTPKEPEPTTPKEPEPTTPKEPEPTTPKEPEPT--------------------    512

Qy   419  PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTPEEPAPTTPKAAAPNTP  478
          |||||||||||||||||||||||||||:|:||:|||||||||||||||||||||| ||
Db   513  -----------------------------------------------------TP      514

Qy   479  KEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT  538
          |||||||||||||:||||||||||||||||||:|||||||||||||||||||||||||:|
Db   515  KEPEPTTPKEPEPTTPKEP------------------------KEPEPTTPKKP---      549

Qy   539  TKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKP  598
          |||||:||||||||||:|||||||||||||||||||| |||||||||||||||||||:||
Db   550  ---------------------------EPTTPKEPVPTTPKEPEPTTP-------KEPEPTTPKEP  581

Qy   599  ARKELAPTTPKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEV  658
          | |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   582  EP----------TTPKEPEPTTPKEPAPTTPKEPEPTTPKEPEPTT----------      617

Qy   659  SPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKIPAATKPEMTTPAKD  718
          |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db   618  --------TSPKT--------------------------------------------TTLK  626

Qy   719  KTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLK  778
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   618  -----------------------------------------------------------  617

Qy   779  TTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKAT-----TPKP-QKPTKAP  829
          ||| |||||||||||||||||||||||||||||||||||     |||| :|||:||
Db   627  ATTLAPKVTAPAE-----EIQNKPEETTPASEDSDDSKTTLKPQKPTKAPKTKPTKAP  681

Qy   830  KKPTSTKKPKIMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLV  889
          ||||||||||:|:|||||||||||||||||||||||||||||||||||||||||||||
Db   682  KKPTSTKKPKI-PKTRKPKTTPAPLKTTSAPPELNTTP---LEVMLPTTTIPKQTPNPETA  738

Qy   890  EVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNG  949
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   739  EVNPDHEDADGGEGEKP-LIPGPPVLFPTAIPGTDLLLAGRLNRGININPMPSDETNLCNG  797

Qy   950  KPVDGLTTLRNGTLVAFRG  968
          ||||||||||||||||||||
Db   798  KPVDGLTTLRNGTLVAFRG  816
```

RESULT 4

Q9N4S7
ID   Q9N4S7;   PRELIMINARY;   PRT;   1079 AA.
AC   Q9N4S7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y51B11A.1 PROTEIN.

```
GN   Y51B11A.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RC   STRAIN=BRISTOL N2;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Cotton M.;
RT   "The sequence of C. elegans cosmid Y51B11A.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match      18.8%;  Score 967;  DB 5;  Length 1079;
Best Local Similarity   32.3%;  Pred. No. 2.9e-54;
Matches  294;  Conservative  114;  Mismatches 377;  Indels 124;  Gaps 33;

Qy   35 TTPDPSTTQHNKVSTSP-KITTAKPINPRPSLPPNSDTS--KEFSLTVNKEITVETKETT 91
         ||:||       |:  :    |||:|::|        |::|    : :|:   ||  :| |
Db  102 TAPETTSTEAPSSSTTPVQTTTTTTAPETTSTEPPSSSTSPVQTTTTTTAPETTSTEAPSSS 161

Qy   92 TTNKQTSTDGKEKTTSAKETQSIEKTSAKDLLAPTSKVLAKPTPKAETTTKGP--ALTTPK 149
         ||:|          ::      | :  |  |   ::|     ||:|:| |       |:|||
Db  162 TTPVQTT------TTTAPETTSTEPPSSS--TSPVQTTTTTAPETTSTEPPSSSTTPV 211

Qy  150 EPTPTPTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTT 209
         | ||| ||||||:|||||| | | || ||||||| |||| || |||||| || |||||| ||
Db  212 QTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPE---TT 268

Qy  210 PKEPAPTTTKEPAPTTTKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPTPTTPKEPA 269
         | |||| ||||||||||| |||| || | |||| ||||||||| || | || |||:
Db  269 STEP-PSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEP-PSSSTTPV 326

Qy  270 PVTTKEPAPTTTPKEPAPTTTAPKKPAPTTTPKEPAPTTTPKEP---------- 315
         | || |||| ||||||||| | | | || || |||||||| |||  :
Db  327 QTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQNTTTTTTAPETTSTE 386

Qy  316 SPTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKKPAPTTAPKEP----APTTPKEP 372
          | ||| |||| ||||||||| || |||| || | |||  |||      || ||||
Db  387 PPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPP 446

Qy  373 TPKKPAPTTPKEPAPTTPKEPAPTTTPKKPAPTTAPKEP------TAP------ 419
         | | |||| |||||||| || ||| || | |||  |||        ||
Db  447 T----APETTSTEPPSSSTTPVQTTTTTTAPETTSTEPP 502

Qy  420 ---------TTPKKLTPTTPEKLAPTTPEKPAP--TTPEELAPTTPEEPPTTPEE---- 468
                  | | |  |||  |:  | ||  ||    || ||  |||| || |||
Db  503 PSSSTTPVQTTTTT----APETTSTESPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQT 558

Qy  469 TPKKAAAPNTPKEP--APTTPKEPAPTTTPKEPAPTTPKGTAPTTLKEPAPTTP 526
         |  |    | ||||  |||| |||||| || ||||| | |||| | |||||| |
Db  559 TTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSS--TTPVQTTTTTTAPETTSTEP 617

Qy  527 KKPAPKELAPTTTKEPTSTTSDKP-APTTPKGTAPTTPKEPAPTTTPKEP--APTTPKGTA 583
          | |||: |||| |||:| | |:| ||||| ||||| |||||| ||||||  ||||| |||
Db  618 PSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTT 674

Qy  584 PTTLKEPAPTTPKPAPKELAPTTTKGPFSTTSDKP-APTTKETAPTTPKEPAPTTPKK 642
         | |||||||| | | ||: |||| |  :|| |:|| ||| |  |||| |||||| ||
Db  675 TTTTAPETTSTEPPSSSSNTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEP- 732

Qy  643 PAPTTPETTPPTTTSEVSTPTTTTKEPTTIHKSPDESTPELSAEPT---PKALENSPKEPGV 699
         |||| |  :  ::|::  :||:|:|||  || ||| |||| |:||     ||| |||| |
Db  733 --PSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSNTP---V 786

Qy  700 PVTTKPAATKPEMTTTAAKDKTTTERDLRTT----PETTTAAPKMTKETATTTEKTTESKIT 755
         | ||:|||||  ::| ||| |||||| | |     ||| |||| | | | || | || |:|
Db  787 QTTTT----TAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSSNTPVQTTTTTTAPETT 843

Qy  756 ATTTQVTRSTTTQDTTPFKITTLKTTTLAPKVTTTK----KTITTTEIMNKPEPTA-KPK 809
         | ||:| :|||:| | |  |||:| || || ||| |     | |||| |||| ||  |||
Db  844 STEPPSSSSTTSPVQT----------TTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPP 894

Qy  810 DRATNSKRATTTPKPQKPITKAPKKPTSTKKPIKTMPRVRKPKTTPTPRKMTSTMPELNPTSRI 869
          ||:|  | ||| | ||  || | | ||| | ||| | |||| ||| ||| :| |||| | |
Db  895 SSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTP------- 946

Qy  870 AEAMLQTTT 878
             : |||||
Db  947 ----VQTTT 951
```

RESULT 5

Q917S1

```
ID   Q917S1      PRELIMINARY;  PRT;  1049 AA.
AC   Q917S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
```

```
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
DR  EMBL; AE003495; AAG22353.1; -.
DR  FlyBase; FBgn0030561; CG5228.
SQ  SEQUENCE   1049 AA;  107278 MW;  954DD629E7619671 CRC64;
```

```
Query Match             18.2%;  Score 936.5;  DB 5;  Length 1049;
Best Local Similarity   34.6%;  Pred. No. 2.5e-52;
Matches  371;  Conservative  45;  Mismatches 439;  Indels 217;  Gaps 54;

Qy      2  RDNKKNRTKKKPFTKPPVVDEAGSGLDNGDFKVTPDTS---TTQHNKVSTSPKITTAKP  58
              ||:|                                |||    ||  || |||| |
Db     30  KPNEIGTTAKPTTLKP-----------------------TEGTSAKPTTLKPTEGTSAKPTTLKP  71

Qy     59  -----INPRPSLPPNSDTSKETSL-----TVNKETTVETKETTTTNKQT--STDGKEKTT  106
                 |:|              |||:|      | ||:|        |:|
Db     72  TEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG-----T  127

Qy    107  SAKET--QSIEKTSAR--DLAPTSKVIAKPT--------PKAETTTKGPALTTPKEPTPTT  155
              |||   ||  |||   |||| |||||    |||  |||||||
Db    128  SAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLK  187

Qy    156  PKE------PASTTPKEPT---PTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAP  207
            |||         |  |||   |||:|||||||||  |||||  ||||||||||  |||||
Db    188  PTEGTSAKPTTLKPTEGTSAKPTTLK--PTEGTSAKPTTLKPTEG                 243

Qy    208  TTTPKEPAPTTPKEPAPTTTKSAPTTPKE---PAPTTPKKEPAPTTPKEPAPTTTKEPT-  262
            |||                                |||||  |||||||  |  |  |
Db    244  TTAK---PTTLKPTEGTTAKPTTLNPTEGTSAKPTTLKPTEGTSAKP               300

Qy    263  TTPKEPAPTTPKEPAPTTPKEPAPTTAPKKKPAPTTPKEPAPTTPK-----EPAPTTTKEPSP  317
            |||                                              |  |||||||:|
Db    301  TTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEG---T  360

Qy    318  TTPKEPAPTTP-------TKSAPTTTKEPAPTTTKSA---PTTPKEPSPTTPKEPAPTTPKE  368
            |||                |:|||||||||   ||||   |||  ||||||||||
Db    361  TTAK---PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEG-----T  416

Qy    369  PAPTTPKEPAPTTPKEPAPTTPKEPAPTTKKPA---PTAPKEPAPTTPKETAPTT--PK  423
            |||                                |||||||||||||  ||||  ||
Db    417  --PTTLKPTEGTTAK---PTTLK-PTEGTTAKPTTLKPTEGTSAKPTTLKPTEG      470

Qy    424  KLTPT---------TPEKLAPTTPEKPAPTTPEELAPT---TPEEPTPTTPEEPAPTTPKAAAPN  476
            |||             |||||||                |||||  |||||
Db    471  TTLKPTEGTTAKPTTLKPTEGTKGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTI  530

Qy    477  TPKEPAPTTPKEPAPTTPKEPAPTTTPKEPAPTTPKGTA---------PTTLKEPAPTTP  526
            |||                                        |||  |||||||
Db    531  KGTTAKPTTLK----PTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTTA  587

Qy    527  K-----KPA-------PKELAPT----------TTKEPTSTTSDKPA                558
            |      |||        |||||          |||  |||
Db    588  KPTTLKPTEGTSAQPTTLKPTEGTSAQPTTLKPTEGTSAQPTTLKPT              647

Qy    559  APTTPKEPAPTTPKEPAPTTPKGTAPTTLKPTEGTTAK-----KPAPKELAPTTTKGPTST  614
            |||                                        ||||
Db    648  EGTTAK----PTTLK---PTTEGTSAKPTTLKPTDGTTAKPTTLNPTEG            701

Qy    615  TSDKPAPTTPKE--TAPTTPKEPAPTTPKKPA---PTTPEPTPPTTSEVSTPTT-TTKEPT  668
            |||              ||||||||        |||
Db    702  TSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTL  761

Qy    669  TIHKSPDESTPELSAEPTP-----------------KALENSPKEPGV--PTTKH---PAATK  709
            ||                                  |          |||||  ||||
Db    762  TL--KPTEGT---SAKPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPTERTSAQPTTLK  816

Qy    710  PEMTTTAKDKTTERDLRTT---------------PETTTAAPKMTKETATT---------TEKTTES  752
            |  |  |        :|                |  ||  | ||  ||         ||||||
Db    817  PTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTTAK  876

Qy    753  KITATTTQVTSTTTQDTTPFKITTLKTTLAPKVTTTKK--TITTTEIMNKPEETAKPKD  810
            |                  |||| ||  |||      ||  |||:|
Db    877  PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLKPTEGTSAKPTTLKPTE  936

Qy    811  RATNSKATTPKPQKPTKAPK---KPT--STKKPKTMPRVRKPKTTPTT---PRKMTSTMP-  861
            |                      |||                          |  ||||
Db    937  -GTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTKGTSGQANNFETKRKERRP  995

Qy    862  ELNPTSRIAEAMLQTTTRPNQTTPNSKLVEVNPKSEDAGGAEG-ETPHMLLRP        912
            |                                                    |:|
Db    996  TLKPTERTSAQ--PTTLKPTEGTTAKPTTLKPTKGTSGQANNFETKRKERRP  1045
```

```
RESULT  6

ID   076894
AC   076894;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:5667.1 PROTEIN.
GN   EG:5667.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
```

PRELIMINARY;   PRT;   1795 AA.

```
RL   Science 287:2185-2195 (2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RP   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003421; AAF45644.1; -.
DR   EMBL; AL031028; CAA19845.2; -.
DR   FlyBase; FBgn0025390; EG:5G7.1.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF01607; Chitin_bind_2; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00494; ChtBD2; 2.
SQ   SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match    16.8%;  Score 865.5;  DB 5;  Length 1795;
Best Local Similarity   28.5%;  Pred. No. 1.4e-47;
Matches 276;  Conservative 129;  Mismatches 447;  Indels 115;  Gaps 28;

Qy      6  KNRTKKKPTPKPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL   65
              ||||     :|  |       |:|:||  |:|      || |:|:  |:|   |  :|
Db    431  KNDVTEAPELIKSPL----KGLHLSENIVILPETTT--TTTTTKPVVLTCPIISPPDTT    483

Qy     66  PPNSDT-------SKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSI      114
              |  |         |:|||        |||    |:|| | |:| | |:|  ||
Db    484  PKPSTTTAVRKSTPKISSTEQHSTTTAKTTTTTKRPTTVTEKTSSATEKPRTTVVTTTTQK   543

Qy    115  EKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPPTTTKS     174
              |:| |   |:|      | |  :|||  |  ||| | |:||||| | ||||  :||
Db    544  RSTTTHNTSPDTIKTTIRSTTLSPKTTTPSTTTPSTTTPST------TTPSTTTPST---   594

Qy    175  APTTPKEPAPTTTKSAPTTPKEPAPTTTK-EPAPTTPKEPAPTTTKEPAPTTTTKSAPTTP   233
              |:|| |  |:||   |   | |  |:|   | ||| | | |: |  | |:|||:||||:|
Db    595  --TTPSTTTPSTTTVKVSTHRPRTTSQKTTTASTTTKK----TTTTSPKTTKTTDIPTST    648

Qy    234  KEPAPTTPKKPAPTTPKEPAPTTTPKEPTPTTPKEPAPTTTKEPAPTTTKEPAPTAPKKPAP  293
              | |  |: ||   :|  |  |||  | | |:|| | |:| | | |: | | |:| ||  |
Db    649  TSKLSTTTQKTTTTTHKFTAAPTTSTEKPKTTTEKTSTVSTTTVKTKSTESSSPKPTSSTGKPT 708

Qy    294  TTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPKEP   353
              ||:|  |  |: | | |:| | |:|:||| | |:|  ||| || | |:|| ||||:| |
Db    709  TTPKPSTRTTTPTTTKVTTTQITTTTPLRSSTETTTSTQPPTTTTTPQPTTTTTLIVTPKTS   768

Qy    354  SPTTTKEPAPTTPKEPAPTTPK--------KPAPTTPKEPAP-----------TTPKE      392
              ||||:||     | |:|| ||          | ||| |: |           ||||
Db    769  TTTTTTTTEKPITSSPKPTTTTQKTTTSTAPNTTKVAITTQKETTPTQSTSTTIFTRKTTTNN  828

Qy    393  PAPTTTPKKPAPTAPKEPAPTTPK-ETAPTTPKKLIPTTPEKLAPTTPEKP--------AP   443
              | |:| || | || |:| |:||  | |||| ||   |||:|  ||| ||          |
Db    829  PEPTSTEKPIPTSTTPKESTTTPKTSVASSTEKTTISSSPKPTTEKSTENPTTNSVKTSAL   888

Qy    444  TTPEELAPTTPEEPTPTTPTPKEPAPTTPKAAAPNTPKEPAPT---TPKEPAPTT       501
              |||:  |||| |:| |:|| ||| | ||| |   ||| |:|||   |||| ||:
Db    889  TSSTQRATSTISEPTKTT-QNITTTTPKPTTLKTSTQEATTSTQKVSTVTITTKAYESS    947

Qy    502  PKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPT--STTSDKPAPTTPKGTA   559
              |  :|||||   |||   |  ||:|  |||||||| :||||   ||| ||| |||||| |
Db    948  PLITTLSTEEPNTTKPLRTTTTTTSVTATTRITTTISESSTETTSTQKPKSTTTPTSTT    1007

Qy    560  PTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPARKELAP-TTTKGPTSTTSDK   618
              |||| | |:|| | |||||||  |||  |  |||| || |  |:|   ||| |||||||
Db   1008  RTTPKVTTVIVSTQNPTTTTSKTSTVTI-----TTP-NPSPSTQRPTTTTTTRQPTSIYAST  1061

Qy    619  PAPTTPKETAPTTPKKPAPTTTPEPPPTTTSEVSTPPTT-TKE----------A       666
              | |:|| | |:|| || |  | |  ||||:  |  ||:| |||          |
Db   1062  TSIGTTRIPTTINPQNSTSSTDLTTVIRPPCDPDSISDKNTNTACTQELQQVNLLELQS    1121

Qy    667  --------------PTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKPPAATKPEMTTTAKD  718
                            |:| |:|| |:|  | | ||| |:|||| | |||| | | ||:||||||
Db   1122  PQKQEQFTHTRTHTALTGSRNTLGGQGEVPDVMDDAPSSAEAESGQANTARAPTMSTLAAA      1181

Qy    719  KTTER---DLRTTPEFTTAAPKMTKETATTTEKTTESKIJTATTTQVTSTTTQDTTPFKIT    775
              |:|:    |  |:|  ||  |||| :|| || |:||  || |::| |:||||| ||| :|
Db   1182  HLLQKLFHIISTTPPSREHAP--TQRPSSQPSSSQRSR-GVTIAQMARHNLATSKPFIAH    1238

Qy    776  TLKTT-----TLAPKVTTTKKTITTTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKA-PK   830
              |||:     ||||:|||:| |:| | |:|||| |:|| | | | ||||||| | |:|  ||
Db   1239  SLRLSIQQLASTQKRSIPPKILVTHNTTKEPED-SEXYDSETSEQYTDEDNEVLDKTQPR    1297

Qy    831  KPTSTKRPKTMPRVRPKPKTTPTPPRKMTISTMPELNPTSRIAERAMLQTTTRPNQTPNSKLVE  890
              || ||| |||| |  ||| |||| |:|  ||  ||:|||| | || ||||||| | ||||:||
Db   1298  AMSSTTVAAVLPAVPSTTTEREPQK-TSSSP--SPT----KATSSTTTQPIETTTGDLEY    1350

Qy    891  VNPKSED                                                        897
              | |:||
Db   1351  DSSGSSD                                                       1357
```

```
RESULT   7
ID   O96449
AC   O96449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RACE 1-11;
RA   Goernhardt B.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061185; AAC72308.1; -.
SQ   SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

Query Match    16.5%;  Score 853;  DB 10;  Length 1489;
Best Local Similarity   32.3%;  Pred. No. 7.3e-47;
Matches 355;  Conservative 46;  Mismatches 475;  Indels 222;  Gaps 49;

Qy      9  TKKKPTPKPKPPVVDEAGSGLDNGDFKVTTPDTSTT-----QHNKVSTSPKITTAKPINPR    62
              | |:|  ||   |  :|:             |:|||      |:||  |:| ||| ||:|
Db    338  TPYAPTEKPYDVEETTYVTEESTYAPTKSETNAPTERMHYAHIEKPCDYEVTMYAPTEET    397

Qy     63  PSLP--------PNSDT----SKETSLIVNKETTVETKETTT---TNKQTSTD         100
              |:|         |:|      ||||      |||  |:|:||   |:| ||
Db    398  TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYPEPTEETTYTEETTYAPTEETTYAP    457

Qy    101  GKEKTTSAKETQSIEKTSAKDLAPISKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPA   160
              | |:|:|||:| | | | |  || | | ||||   |||| | |  |||| | ||||| |
Db    458  TEETTYAPTEETPYEPTEETTYAPTEKTTYAPT---EETTYASTEETTYAPTEETTYAPA    514

Qy    161  STTTPKEPTPTTIKSAPTTPKEPAPT-TTKSAPTTPKEPAPT--TTKEPAPTTPKEPAPTT  217
              | || | | | | ||| ||||| |||  |:| |:||||| |   ||| |||| | | |||
Db    515  EETPYEPTEET-TYAPTEEPTEETTYAPTEETTYAPTEETTYAPTEEPTEET          573

Qy    218  TKEPAPTTTKSAPT-----------TPKEPAPTTPKKPAPTTPKEPAPTTPKEPPTTPKE   267
              | | |:|  ||||            || | |||| || || || | |:||| ||  |||
Db    574  TYAPTEETT-YAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYA    632

Qy    268  P----------AP--TTKEPAPTTPKEP-----APTTAPKKPAPTTPKEPAPTTP       304
              |          |   |:| |||| ||      |||| || || || | |||| |
Db    633  PTEETTYASTEETTYAPTEETTYAPAEETTYEPTEETTYAPTEETTYAPTEETTYAPTEE    692

Qy    305  KEPAPT---TTKEPSPTTPKEP-----APT-TTKSAPT-----------TYKEP-----A  339
              | | ||   |:| | | |||      || | |:||||           |:|||     |
```

```
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR001998; Xylose_isom.
DR   InterPro; IPR002955; P_rich_extensn.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS001172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR   SMART; SM00370; LRR; 3.
SQ   SEQUENCE  1315 AA;  134401 MW;  64C97A2A01F09336F CRC64;

RESULT 8
Q9SPM0
ID   Q9SPM0       PRELIMINARY;       PRT;   1315 AA.
AC   Q9SPM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN.
DE   PEX2.
GN   SEQUENCE FROM N.A.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA   Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT   "Pollen Extension-like (Pex) Genes in a Monocot and a Dicot.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159297; AAD55980.1; -.
```

```
Query Match             15.7%;
Best Local Similarity   25.0%;          Score 808.5;   DB 10;   Length 1315;
Matches  255;  Conservative  88;   Pred. No. 4.6e-44;
                                    Mismatches  459;   Indels  219;   Gaps  33;
```

```
Qy   693 TTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPIEETTYGPTEETTYA 752
             |||   ||   |: |        ::|    |||
Db   340 PT--TTKSAPT--TPKEPSPTTTKEP-----------APTTPKEPAPTTPKKPAPTTPKEPAPT-- 388

Qy   753 PTEATTYAPTEETPYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEE 812
             |||  |||       |||       |||
Db   389 TPKEPAPTTKPAPTAPKEPAPTTPKKLTPTTPEKLAPTTPKRPAPTTPEE 448

Qy   813 TPYEPTEETTYAPTEETPYEPTEET--TYTPTEETTYAPTEETTYAPTEETT 870
             |||          :||
Db   449 LAPT--TPEEPTPTTPEEP-------APT----TPKAAAPNTPKEPAPTTPKEP- 489

Qy   871 YAPTEETPYEPTEETTYAPAEETPYEPTEETTYAPIEETTYAPAEEETPYEPT 930
             |||        |||
Db   490 ----APTTPKEPAPTTPKETAPTTPKKPAPK-ELAPT--TT 539

Qy   931 EETTYAPTEETPYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 990
             |||           |||
Db   540 KEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP-----APTTPKGTAPT 585

Qy   991 YAPTEETMYAPIEETTYAPTEETTYAPTEETTYAPTEETTYAST 1050
             |||
Db   586 ---------TLKEPAPTTPKKP------APKE--LAPT--TTKGPTSTTSDKPAPTTP 624

Qy  1051 EETTYAPTEETTYAPAEEETYAPTEETTYASTEEETTYAPAEEETPYEPT 1110
             |||
Db   625 KETAPTTPKEPAPTTPKEPTPETPPPTTSEVSTPTTTPKEP--TIHKSPDESTPELS 682

Qy  1111 YAPAEETPYEPEETTYAPTEETMYAPIEETTYGPTEETTY-APTEATTYAP 1169
             |||
Db   683 AEPTPKA-LENSPKEPG------VPTTKTPAATKPEMTTTAAKDKTTERDLRTTP---ETT 732

Qy  1170 TEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETPYEPTEETT 1229
             |||
Db   733 TAAARKMTKETATTTEKTTESKIITATTTQVTSTTTQDTT------PFKITT--LKT 779

Qy  1230 TYAP-------TEETTYEPTEETTYAPTEETTYAPTEETTYGPTEE 1284
             |||
Db   780 TTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAT--NSKATTPKP-QKPTKAPKKPT--- 833

Qy  1285 TTYAPTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPMEETPYE 1344
             |||
Db   834 ---STKKP-KTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP--N 881

Qy  1345 PAEESTSTVSTEKPCNTEEFTDEPTDEPTDE--PSDEPTDEPTDLPTDEPSTPCDN 1402
             |||
Db   882 STKKP--KTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP--N 899

Qy   882 QTPNSKLVEVNPKSEDAG 899

Qy  1403 QGINGIGVENKVRNNAG 1420
```

```
Qy    13 PTP------------------KPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPR 62
             |||                ::|                   ::|      ::|   :|  |
Db   412 PTPGGPPSSSVPGKPPSVPG-------KPAAPAPMPTPHTPPDVSP--EPL-PE 455

Qy    63 PSLPP------NSDQSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQ 112
             |||
Db   456 PSPVPAPAMRMPLRSPPADEYIPT---PPVPAKSPPGTSPPASRGAPPLQOAPPAAS 511

Qy   113 SIEKTSAKDLAPTSKV-----------LAKPTPKAETTTKGPALITPKEPTPTTPKE 158
             |||
Db   512 SPPATPVKSSPPPAAVVLPPPAKTPSPPAPVASPPPEAPVSSPPQPVKSSPPPAPVASPP 571

Qy   159 PASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTP-KEPAP----TTTKEPAPTTP 210
             |||
Db   572 PPMKSP--PPPARVASPPPLMKSPPPPAPVASPPPQPLKSPPPPVLWLSTPSVKSPPPPVP 629

Qy   211 KEPAPTTKEPAPTTPKSAPTTPKE---PAP--TTP----KPRAPTTPKEPAPTTPKE- 259
             |||
Db   630 VASPPPVKSPPPLAPVSSPPPPVKLPPLPAPGKSTPPPEEEKPTPPPPVKSSPPPPEKSL 689

Qy   260 PTPTPTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPSPTT 319
             |||
Db   690 PPPLITTSPPPQEKPTPPSTPSKPPPPSPVETLPPPSKSSSPPEEPVSSPPPQPVSSP-- 747

Qy   320 PKEPAPTTTKSAPTTTKEPAPTTTKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 379
             |||
Db   748 ---PAPVS--SPPLKSSPPPQVEKTSPPPLAPVSSPPPTPKSSSPPLAPVSSPPPAPV 803

Qy   380 TTPKEPAPTTPKEPAPTTTKKPEPSPTTTKKPEPAPTTPKKLLTPTTPEKLAPTTPE 439
             |||
Db   803 PTPKSSSPPLAPVSSPPQVEKTSPPPAPVSSPPPTPKSSSPPLAPVSSPPPQVEKTSPPPAPV 862

Qy   440 KPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAP 499
             |||
Db   863 SSPPLEPKPSSPPSVSSPPTTVKSSPPPMTPKSSSPPAHVSSPPPEAEKSSPP 922

Qy   500 TTPKETAPTTPKGTA------------PTTLKEPAPTTPKKPAPKELAPTTPKEPISTTSD 548
             |||
Db   923 LAPISSPPSEPKSPSSPSSPPMVEKTSPPPATVSSPPPVKSSPPP--APVSSPPPVVKSSP 979

Qy   549 KPAPT-----TPKGTAPTTP------KEPAPTTPKEPAPTTPK--GTAPTTLKEPAPT 593
             |||
Db   980 PPAPVSSPPPTPKPLPPPAPVSSPPPTVKSSPPPPTPKPLPPPPPTPVSSPPPT 1039

Qy   594 TPKKPAPKELAPTTTKGPTSTTSDKPAPTT--PKETAPTTPK-----EPAPTTPKKPAPTT 647
             |||
Db  1040 XKPLPPP---APVSSPPPVVKSSPPPAPVSLPPPRKPSSPPRTRVSSPPPVVKCCPPPTL 1096

Qy   648 PETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA 707
             |||
Db  1097 VSSPPPAPKSLPPPPPVSSP---------PPEVKSSPPPTPVSSPPPAPKSSSPPPTPVS 1146

Qy   708 TKPEMTTTAKDKTTERDLRTTP-----ETTTAAPKMTKETATTTEKTTESKIITATTTQVT 762
             |||
Db  1147 SPPP----------ELKSSPPPAPVSSPPSAPKSSPPAPVSLPPPPEVKSSPPPPAPIS 1194

Qy   763 STTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKP 822
```

```
Db  1195 SPPPPAKSP----------------PPPA-----------PMSSLPPPVKSPPPPAPVSSP 1228

Qy   823 QKPTKAPKKPTSTKKKPTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQ 882
              ||:|                            |:|                 |  |
Db  1229 PPPMKSPPPPAPISSPPPAP--VKPPSLPPPAPVSSSPPPAV----TSAP-- 1271

Qy   883 TPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGTIINPMLSD 942
                ||:|     ||||:|     |:||                |||:  |: ||::::
Db  1272 ----------PKKEEDSTA-------------PPAEALPPPSFNDIILPPIMAN 1302

Qy   943 E 943
         :
Db  1303 K 1303

RESULT   9
ID   Q20007
AC   Q20007;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   COSMID F35A5.
GN   F35A5.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RL   Nature 368:32-38(1994).
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Leimbach D.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U46675; AAB52641.1; -.
DR   EMBL; U46675; AAB52641.1; -.
SQ   SEQUENCE 1274 AA; 138065 MW; 07F6BOD292C1799F CRC64;

Query Match       15.6%;    Score 802;  DB 5;    Length 1274;
Best Local Similarity  28.0%;   Pred. No. 1.2e-43;
Matches 290;    Conservative 102;  Mismatches 414;  Indels 230;  Gaps 56;

Qy     6 KNRTKK-KP-------TP-----KPPVVDEAGSGILDNGD----FKVTTTPDTSTTQHNKVSTS 50
           ||                   |:|                 || |
Db   274 KNPTKKWKPPWEDDETPVEEVKEPPVPEKKAPVLKKKDPAPAAKARDPSPSKAAPKKVEPS 333

Qy    51 PKITTAKPI-NPRPSLPPPNSDTSKETSLTIVNKETTVETKETTTTNKQTSTDGKEKTTSAK 109
```

```
Db   334 SPVVPPTPVKNPVKKYKPPWEDDEPAEEVKKPSAPEKKTPVLKRKEPEPSSTTPSSDPS 393
             ::  ||                                  ||

Qy   110 ETQSIEKTSAKDLAPTSKVLAKPYPKAE---------TTTKGPA 148
         | |:|                                |: |
Db   394 PKKAAPAVKPRDSSPKKATIPLQADPKAQEVPPTPVKNPVKKYKPPWEVDDEDPVEEVKQP 453

Qy   149 KEPTPTTP---KEPASTTPKEP---TP-TTIKSAPTTPKEPAPTTTKSAPTTPKEPAP 199
         :||  |        ||||||    | :|   ||
Db   454 EAPAKKTPVLKRKEPAAKDTAKPAISKTPETPEKKDPVKRPDSSPKKVAAKPDSAQAPA- 512

Qy   200 TTTKEPA---------------------PTTKEPAPTTTKEPAPT----TTKSAPTTP 233
             |:                  |:|           ||
Db   513 TPVKNPVKKWRPPWEDDEDETPADDVSKPTDAKKTPSLAKKDPAPAKESLKPKADTKAPAKP 572

Qy   234 KEP--------APTTPKKPAPTTPKE---PAPTTPKEPTPTTPKEPAPTTTKEPAPTTPK-EP 283
         :|:              ||           |::|: |:  ||
Db   573 RDPSPKKVAPTAPEKKTPVLAKKEPAGPADSKTKEPEKSKPRDPSPKKAVPAKPVPKTEV 632

Qy   284 APTTAPKKPAP-----TTTPKEPAPTTPKEPAPTTTKEP----SPTTPKEPAPTTTKSAPT 333
         |||      |        ||||||  ||         :|     | ||||
Db   633 APAAVKKPEPISKPKDTAPKKAEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPE 691

Qy   334 TTTKEPAPTTTKSAPTTP-KEPSPTTTKEPAPTTPK--EPAPTTPKKPAP-TTPKEPAPT 388
             ||||||||    |         ||              ||       |||
Db   692 PEKK-TPVLAKKAPTKPDSEAAADPVSGPSSKDPKLAKKAAPVKPRDSSPMKAVPIKPAPK 750

Qy   389 TPKEPAPTTTKKPAPTA-PKEPAPTTPKETAPTTPKKLTPTTP----EKLAPTTPEKPA 442
             |||||||:: |: |:|:|::|||:|| |||::|||||            ||||
Db   751 T--EVPPAVVKKPEPVAKSRDPSPKKAK-AEPNSP--VVPPTPVKNPVKKWKPPWEDDA 805

Qy   443 PTTPEELAPTTPEEP-----TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTP-PKE 496
                       |:|:          ||   ::||
Db   806 PAEPVNVPEPEKKTPVLAKKTPVKPRDPSPKKAVPAKPSTKIDAPPVSVKPEPVSKPKE 865

Qy   497 PAPTTPKETAPTTPKGTAPTTTPKKPAPKELAPT-TTKEPTSTTSDKPAPTTP 555
             |: |    ||                 ||||:  |::||
Db   866 PSPKKAEPNSPVVP---PTPVKNPVKKW-KPPWEDDDDEPTEEVKKPSE--PEKKTPVLA 918

Qy   556 KGTAPTTTPKEPAPTTPKEPAPTTPKGTAPTTLTKEPAPT---------TPKKPAP----KEL 603
         |:|           ||       |: |              |        :|:|||
Db   919 K-KEPEKPKD-APKVAAKPRDPSPKKAVPE---KEPAKVAAKPRDLSPKKAIPIPANTQEA 974

Qy   604 APTTTKGPTS----------TTSDKP----------APTTPKET----APTTPKEPAPTTTKKPAPT 646
             |: |||:           |::||             ||:|    |:|| ||||
Db   975 PPTPVKNPVKKWKPPWEDDDEPAEPVSAPEPEKKTPVLAKKAPAKPRDP---SPKKAAPV 1031

Qy   647 TPETTPPTTSEVSTPTTTKEPTTIHKSP---------DESTPELSA-EP----TPKALENSPKEPG 698
                  |: |: |:|                      ::|        ||||
Db  1032 AAK-PDPKIPEV--PPTPVKNPVKKWKPPWEDDDDEPSEPVSAPEPEKKTPVLAKKAPTKPA 1089

Qy   699 V-----------------KPEMTTAKDKTERDLRTTPETT--TAAPK 737
                             ||  |:||||: |: |
Db  1090 TKPDSEAAADPVSGPTSKDPKLSKKAPAVEKPKPTTDPKDDKLRPSPAAKPEKAPEPAAPK 1149

Qy   738 MTKEIATTTEKTTTESKITATTTQVTSTTTQDTTTPEKITTILKTTTLLAPKVTTTKKTTTTTTE 797
         | |   : || | :         ||:              ||   ||
Db  1150 KWKPVWDDDPDDEPEADTTVPAPSKKPDTEDPADPLG------GPKTKDPK-------- 1193

Qy   798 IMNKPEETAKPKDRATNSKATTTPKPQKPTKAPKKPTSTKKPKTMPRVRKPK----- 848
          | |||:||    :| ||| ||  ||:|||| |    |::|                |
Db  1194 -LNKKAPAEKPTEK--------PKPKEVSKEPPKPTEPPKP-AAPKKWKPPWEDDPDEPE 1243

Qy   849 ---TTTPPRKMTSTMP 861
            |||
Db  1244 ADFTMPAPKKPDTEDP 1259

RESULT  10
ID   Q9LIE8
AC   Q9LIE8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
```

(further right column)

```
PRELIMINARY;         PRT;  1274 AA.



Q9LIE8              PRELIMINARY;    PRT;  1480 AA.
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   STRAIN=COLUMBIA;
RX   PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
RL   EMBL; AP001306; BAB03062.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE 1480 AA; 147153 MW; D1ACC079E155E732 CRC64;

Query Match       15.5%;   Score 800.5;  DB 10;  Length 1480;
Best Local Similarity  29.6%;   Pred. No. 1.7e-43;
Matches 307;  Conservative 54;  Mismatches 505;  Indels 171;  Gaps 45;

Qy   10 KKKPTPTPKPPVVDEAGSGLDNGDFKVTPDTSTTQHNKVSTS-----PKITTAKPINPRPSL 65
            |:|| ||||| | |||
Db  117 KPRPHPKPPNVK-----------PHPHPKPPTKPHPHPKPPTKHPHPHPKPPTKPKPPSV 167

Qy   66 PPNSDTSKETSLIVNKETTVETKETTTTINKQTSTDGKEKTTSAKETQSIEKTSAKDLAPT 125
           : | |
Db  168 KPPPSTPK------------PPTINPPPSTPQPPTHKPPPGCSTPPPVASPPMATPP 211

Qy  126 SKY--LAKPTPKAETTTKGPALLTPKPEPTP--TTP--KEPASTTPKEEPTPTTIKSAPTTP 179
            :: |
Db  212 TQMPPIATP-PIAKSPVATPPIATPPIATPPITIPPVATPPIATPPIANPPIIMPPIATP 270

Qy  180 KEPAPTTP------KSAPTTPKEPAPTTP------------------KEPAPTTPKEPAPTTP 220
             |||
Db  271 PVAAPPITNPPISKPPVTTP------PTTTPPIAKPPIATPPISTPPAANTPPAATPPIITTLP 326

Qy  221 PA-------PTTTKSAPTTP-KEPAPTTPKKPAP---TTPKEPAPTTPKEPTPTTPKEPAP 270
            |||
Db  327 PAKPPVAISPIVT----PPVITPIAQPPVATPPIATPPVATPPIATPPISKSPISTPPISESP 384

Qy  271 TTKEPAPTTP-KEPAPTTPAPKPPAPTTPKEPAPTTPKEPAPTTP----KEPSPTTPKEPAP 325
             |||
Db  385 VATPPPIATSPIKTPPPAKPPVATPPIATPPIATPPVATPPIEKPPVATPPTTPP 444

Qy  326 TTTTKSAPTTTKEPAPTTTKSAPTTPKEP--SPTTKEPAPTTPKEPAP---TTPKKPAPT 380
             :|||
Db  445 TAT---PPVAKPVETPPIATPPTAKPPISTPPISKPPVANPPAATPPIITTPTPVKPPVA 501

Qy  381 TPKEPAPTTPKEP--APTTPKPAPTAP--KEPAPTTPKETAPTT------PKKLTP---T 428
             |||
Db  502 TPPLAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVATPPIAKPPVVTPPTT 561

Qy  429 TPEKLAPTTPEKPAPTTPEELAPTT-PEEPTPTPEEPAPTTPKAAAPNT---PKEPAPT 484
             |||
Db  562 TPPTAIPPVAKPVATPPIATPPTAKPPISTPPISKSPVATPPAATPPIITTPPPAKPPVA 621

Qy  485 TPKEPAPTTPKEPAPTTPKETAP---TTPKGSTAPTT---LKEPAPTTPKKPAPKELAPTTT 539
             :|:|
Db  622 TPPIATPPIAKPPVATPPIATPPIATSPVAKPVATPPIATPPIKTPPPAKPPVAIPPIATPPVA 681

Qy  540 KEPSTSTTSDKPAPTTPKGTAP-TTPKEPAPTTPKGTAP-TTPKEPAPTTPKGTAPTTLKEPAPTTPKKP 598
```

```
Db  682 KPPVAT----PPTATPPIATPPIATPPVVTPPTATPPVATPPIAKPPTIT--PPTATPPVA 736
             |:|
Qy  599 APKELAPTTPKGPTST--TSDKPAPTTPKETAPT-TPKEPAPTTPKKPAPTTP-----E 649
             |||
Db  737 MPPIATPPPAKPPIATPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVA 796
             |||
Qy  650 TPPPTTSEVSTPTTTKEPTTIH--KSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA 707
             |||
Db  797 TPPIAKPPVATPPTTAPPIAIPPVAKPPVATPPTTATPPPAKPPILTPPISKPPVATPPAA 856
             |||
Qy  708 TKPEMT-TTAKDKTTERDLRTTPETTTTAAPKMTKETATTTEKTTESKITATTTQVTSTTT 766
             :|:|
Db  857 TPPITTPPPAKPPVATPPITATPP--IAKPPVATPPIATPPIATSPVAKPPVAIPPIKTPP 914
             |||
Qy  767 QDTTPFKITTLKTTTLAPKVTTTKKT----ITTTEIMNKPEETTAKPKDRATNSKATTPKP 822
             :|:|
Db  915 PAKPPVAIPPIATPPVAKPPVATPPTATPPIATSPIATPPVVT---PPTATSPVATPPIA 971
             |||
Qy  823 QKPTKAPRKPPTSTKKPKTMPRVKPKPKTTPTPRKMTSTMPELNP------TSRIAE---871
             |||
Db  972 KPPTTTP--PTAT-PPVAMPPIATP--PTAKPPVATPPIANPPVEKPPVATPPIAKPPT 1025
             :|:|
Qy  872 -----AMLQTTTRPNQPTPNSKL--VEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMD 924
             ::|:|
Db 1026 VLPPIAKPPVETSPTATPPTATPPVAIPP--------VKPPVAIPPITKPPV 1070
             |||
Qy  925 YLPRVPNGGIIINPMLS 941
             ::|
Db 1071 ATPPVTNPPTAMPPIVT 1087

RESULT 11

ID   Q41805     PRELIMINARY;   PRT;   1188 AA.
AC   Q41805;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z34465; CAA84230.1; -.
DR   Mendel; 14346; Zeama; 2366; 14346.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003592; LRR.
DR   Pfam; PF00560; LRR; 3.
DR   SMART; SM00370; LRR; 4.
KW   Signal.
FT   SIGNAL    1    27    POTENTIAL.
SQ   SEQUENCE 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

Query Match       15.2%;   Score 784;  DB 10;  Length 1188;
Best Local Similarity  29.9%;   Pred. No. 1.5e-42;
Matches 230;  Conservative 59;  Mismatches 359;  Indels 122;  Gaps 33;

Qy   13 PTPKPPVVDEAGSGLDNGDFKVTPDTSTTQHNKVSTSPKITTAKPINPRPSL----PPN 68
             |||
Db  460 PTPHSPPAD------DYVPPPPVPGCKSPPATSPSPQVQPPAASTPPPSLVKLSPPQ 510
             :|:
Qy   69 SDISKETSLIVNKETTVETKETTTTINKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV 128
             |||
Db  511 APVG-------------------SPPPPVKTTSPPAPIG----SPSPPPVSVV 541
```

```
Qy    129 -----LAKPTPKAETTKGPALTTPKEPTPTTPKEPPTTPKEPASTTPKEPPTTIKSAPTTPKEPA 183
              :||  |    ||:|     |  ||| :  |||  | | ||||  | ||  :|  || |
Db    542 SPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVASPPPPVKSP--PPPTLVASPPPPVKSPP 599

Qy    184 PTTTKSAPTTP--KEPAPTTTKEPAPTTTKEPAPTTTKSAPTTP-KEPAPTT 240
              | ||:| ||    ||||||| ||||||| |||   |:| ||   |||| |
Db    600 PPAPVASPPPPVKSPPPPTPVASPPPPVASSPPPPMKSPPPPTPVSSPPPPEKSPPPPP 659

Qy    241 PKKPAPTTTKEPAPTT-------PKE--PTPTTPKEPAPTTTKEPAPTTPKEPAPTAPKKP 291
              | :|||||||||||       |||  | |||||||||| ||||||||  |   | ||:|
Db    660 PAKSTPPPPEEYPTPPTPISVKSSPPPPEKSLPPPTLIPSPPPPEKPPPPSTPSKP-PSSSEKP 718

Qy    292 APTTPKEPAPTTPKEPAPTTTKEPSPTTTKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPK 351
              | ||||||||||||  | ||  |   |||||  | ||:| |||||||||||||||||
Db    719 SP--PKEPVSSPPQTP---KSSPPPAPVSSPPALAPVSSSPPSVKSSP---- 768

Qy    352 EPSPTTTKEPAPTTPKEPAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTPKTTAPTTKKP 411
              :|  |||||||||||||||||  | | || |||  |  ||| ||| ||   |:|| :||
Db    769 PPAPLSSPPPAPQVKSSPPPVQVSSP--PPARKSSPPLAPVGSSPPPPQVEKTSPPPAPLSSPP 827

Qy    412 TTPKETAP-------TTP---KKLTPTTPEKLAPTTPEKLAP 458
              || |  |       || |    :|  |||  |||   |  |
Db    828 LAPKSSPPHVVVSSPPPVVKSSPPPPAPVSSPPPPLTPKPASPPAHVSSPPPEVVKPSTP--PA 885

Qy    459 PTTPEEPAPTTPKAAANTPKE-PAPTTPKEPAPTTPKETAPTTPKGTIAPTTP 517
              | | | ||||||   | ||  ||||||||||||||| | |||||  ||| |
Db    886 PTTVISP-PSEPKSSPPPTPVSLPPPIVKSSPPPAMVSSP-PMTPKSSPPPVVVSSPPPT 943

Qy    518 LKE---PA------PTTPKK---PAPKELAPTTTKE---PTSTTSDKPAPTTPKGTAPTTP 563
              |||    |      || |:|    ||| |||| |:|   | |||||| |||||| ||| |
Db    944 VKSSPPPPAPVSSPPPPATPKSSPPPPAPVNLPPPEVVKSSPPPPVSSPPA---PKSSPPPAP 1000

Qy    564 -KEPAPTTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTPKGPTSTTSDKPAPT 622
              :||| || ||||| | |  | || |||||||| | : || | |||||| | |||||||||
Db   1001 MSSPPPPEVVKSSPPPPAPVSSPPPPATPKSSPPPPAPVSSPPP---PPVKSPPPPAPVSSPPPV 1057

Qy    623 -TPKETAPTTPKEPAPTTPETPTPTTPSEVSTPTTTKEPTTIHKSPDESTPEL 681
              ||||  ||||||||||||   |   |  :||  |||   | | |||  ||||
Db   1058 KSPPPPAPISSPPPPVKSPPPPAPVS---SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS 1115

Qy    682 SAEPTTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPET 731
              | ||||| |  ||| |  |  ||:| :| |  ||: |:| ||    ||||
Db   1116 SPPPAP-----VKPPSLP-PPAPVSSPPPPVVTPAPPKKEEQSLPPPAES 1158
```

RESULT 12
P70670

ID   P70670   PRELIMINARY;   PRT;   2187 AA.
AC   P70670;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE   MUSCLE-SPECIFIC FORM GP220).
GN   NACA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96312450; PubMed=8698236;
RA   Yotov W.V., St-Arnaud R.;
RT   "Differential splicing-in of a proline-rich exon converts alphaNAC
RT   into a muscle-specific transcription factor.";
RL   Genes Dev. 10:1763-1772(1996).
DR   EMBL; U48364; AAB18734.1; -.
DR   EMBL; U48363; AAB18732.1; -.
DR   MGD; MGI:106095; Naca.
DR   InterPro; IPR002715; NAC.
DR   InterPro; IPR003037; TS-N.
DR   Pfam; PF01849; NAC; 1.
DR   Pfam; PF02044; TS-N; 1.

SQ   SEQUENCE   2187 AA; 220599 MW; 003646AA864DEBFD CRC64;

Query Match   15.1%; Score 776; DB 11; Length 2187;
Best Local Similarity   27.7%; Pred. No. 8.6e-42;
Matches 288; Conservative 122; Mismatches 389; Indels 240; Gaps 49;

```
Qy     34 VTTPDTSTTQHNKVSTTSPKITTAKPINPRSLPPNSDISKE------------TSLTVNKETT 84
              :|      |: :  |  |  || :| |   |||| |||              || ||||
Db    806 VQSPKVDPIMSDVTPTSPKKTSATAV-PK------DTSATLSLKSVPAVTSLLSPPPKAPV 857

Qy     85 VETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPA 144
              | :|::| |  ||  :|:|||:|| ::  |||  :|  | |||   | |     | |||
Db    858 APSNEATIVPTEIPISLKNALAAAIPKEILATSIPKVTSPS---PQKTPKSVSLKGAPA 913

Qy    145 LTT-----PKE----PTPTTPKEPASTT----------PKEPTPTT 171
              |||    |||    | |||||||| ||          ||||| ||
Db    914 MTSKKAATEIAASKDVSPSQOFPKEVPLLQHVPPTSPPKSPVSDTLSGALTSPPPKGP-PAT 972

Qy    172 IKSAPTTPKE-PAPTTTKSAPTTPK-------TPKEPAPTTPKK-PAPTTPKEPAPTTP 221
              |||| ||||   ||||||||||||       ||||||||||||  ||||||||||||
Db    973 LAETPTYKPKKSPKPAASKKTPATPSPEGVTAVPLEIPPCSKKAPKTAAPKESSATSSSKR 1032

Qy    222 APTT---TKSAPT-------------------------TPKEPAPTTPKK-PAPTTPKEPAPTTP 257
              ||||   ||||||                          ||||||||||   ||||||||||||
Db   1033 APKTAVSKEIPSKGVTAVPLEISLPLKETSKSATPGEKSASSPKRSPKTAGPKE---TPP 1089

Qy    258 KEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKP------APTTPKE-PAP----TTPKEPAPTT 311
              ||  | ||||||||||||||||||||||||||   ||||||| | ||    ||||||| ||
Db   1090 GGVTAVPPEISLPPKETPQNATPNESLAASSOKRSPKTSVPKETPPGGVTAMPLEIPSAP 1149

Qy    312 TKEPSPTTPK-----EPAPTTTKSAPTTKEPAPTTPK-------EPSPTT 357
              | |||||||     |||||||||||||||||||||       ||||||
Db   1150 QRAPKTAVPKQIPTPEDAVTLLAGSPLSPKKASKTAAPKEAPATPSVGVIAVSGEISPSP 1209

Qy    358 TKEPAPTTPKEPAPTTPKKPAPTT---PKE------------PAPTTPKEPAP 396
              |||||||||||||||||||||||   |||            ||||||||||
Db   1210 KKTSKTAAPKENSATLPKRSPKTAAPKETPATSSEGVTAVPSEISPSPPIPASKGVPVT 1269

Qy    397 TTPKKPAPTTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEK-PAPTTPEELAPTTP 455
              |||| | ||| ||||||||||  | |||||  ||||||||||||   |||||||||||
Db   1270 LTPKGAPNALAE-SPASPKKVPKTAAPEETSTTP----SPOKIPKVAGPKEASATPPS 1322

Qy    456 EPTPTT-------PEEPAPTTPKAAAPNTPKEPAPTTPKE-PAPTTPKEPAPTTPKETA 506
              |||||       | ||||||||  | ||||||||||||   |||||||||||||| |
Db   1323 KKTPKTAVPKEETSAPSEGVTAVPLEIPPSPRKAPKTAAPKETPAPS--PEGATTAPVQIP 1380

Qy    507 PTTPKGTAPTTLKEPAPTTP---------------------KKPAPKELAPTTTKEPTS-- 544
              |||||||||||||||||||                      |||| ||||||||| |
Db   1381 PSPRKGSKKAGSKE-TPTTPSPEGVTAAPLEIPISSKKTSKMASPKETLVTPSSSKKLSQT 1439

Qy    545 ---TTSDKPAPTTPKGTAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKKPA 599
              ||||||||||||||  ||||| | ||||||||||||||||||| |
Db   1440 VGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-A 1496

Qy    600 PKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETP------PPT 654
              |||||||  | |||||||||  |||||  ||||||||||||||||||        |||
Db   1497 PKTAAPPSER-VTIVPPEKPA-TPOKASGTTASKVPVPAETOEVAVSSRETPVTPAVPV 1554

Qy    655 TSEVSTPTTTK----EPTTIHKSPDESTPELSAEPTPKALENSPKE-PGVPTTKTPAA 707
              ||| ||| |    |||  ||||| |||||  |  ||| | ||||   || |||||
Db   1555 KNPSSHKKISKTIELKEAPATLPPSPTKSSPKIPSSKKAPRT-SAPKEFPASPSIK-PVT 1611

Qy    708 TKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITPATTTQVTSTTTQ 767
              ||| ||||||||||||||||| |  |:| ||| || |    |:| | ||  ||
Db   1612 T--SLAQTAPPSLQKAPSTTIPKENLAPAV-------LPVSSKSPAAPARASASLSPATAAP 1665

Qy    768 DTTPFKITILKTTILAPKVTTTKKTITTTEIMNKPEEFAKPKDRATNSKATTTPKQKPTK 827
              || ||||||||  || |||| |:|||||| |:| ||: | | ||||||||||| |||||
Db   1666 QTAPKEATTLIPSCKKAAATETPIEITSTAPSLEGAPKETSE----TSVSKVLMSSP---- 1716

Qy    828 APKKPTPSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSK 887
              || | :| ||| || |  :|: |||:|  | || ||| ||||  | ||  ||  |||
Db   1666 ... 887
```

```
Db  1717 -PKKASSSKRASTLP------ -ATTLPSLKEASVLS----- -PFATSSGK
Qy   888 LVEVNPKSEDAGGAGECETP 906
Db  1753 DSHISPVS-DACSTGTTTP 1770
```

```
RESULT 13
Q9VEL9
ID   Q9VEL9      PRELIMINARY;      PRT;   2112 AA.
AC   Q9VEL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG4090 PROTEIN.
GN   CG4090.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003716; AAF55402.1; -.
DR   FlyBase; FBgn0038492; CG4090.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR005561; EGF-like.
DR   Pfam; PF01607; Chitin_bind_2; 10.
DR   SMART; SM00494; ChtBD2; 11.
DR   PROSITE; PS00022; EGF_1; 1.
SQ   SEQUENCE 2112 AA; 219547 MW; B91018E551A5D36 CRC64;
```

```
Qy  Query Match 14.3%;  Score 739;  DB 5;  Length 2112;
    Best Local Similarity 27.2%;  Pred. No. 1.9e-39;
    Matches 281;  Conservative 98;  Mismatches 339;  Indels 316;  Gaps 44;

Qy     4 NKKNRETKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRP   63
            :||  |    |:      || :|| :|              |||| |||      |
Db  1023 NHPDQIQTKRPLKCKRVVSQGGSSSN------ -STSNSSSSSNNSGSSSSSNSGSGSSSS-- 1070

Qy    64 SLPPNSDTSKETSLITVNKETTV- -ETKETTTTNKQTS- --TDGKEKTTSAKETQSI--  114
            :||||         |||||||                   |||||
Db  1071 SSSSNSGSSSSNTGSSSNSGASSSGGGSSNGQSSSSNSSGSGSGSSSSTSSSSSSSSS 1130

Qy   115 EKTSAKDLAPTSKVLAKPTPKAETTTKGP------                         143
            : ||
Db  1131 NNNNQGSSSSSSSSSSSSSSSTSSKPNPSETCKVNGQFIGDRSDCAKFYRCVNDNDRGGFNMVPF 1190

Qy   144 ------ALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTT                   178
                     :||              |||
Db  1191 SCGPGTTVWDAQMQACNHAWAVKECGGIAPPTTSTPTTSR-PTTASTSRPSDQTSTSRPTG 1249

Qy   179 PKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTT------               217
              :|||
Db  1250 P------PTTARPVTARPTTSSPTTASSSQTTTSPVTQAPNTDGKCRSEGFMADPNNCSKFYR 1305

Qy   218 ------ -TKEP----APT------ -TTKSAPTTPKEPAPTTPKKRAPTT         248
                  :|||        |:|
Db  1306 CVRNNKGGFTSIPFQCCGAGTVWDDQLQTCNHNFNNCSTGTESTTPKKPC--EPATNGTTA 1363

Qy   249 PKEPAPTTPKEPTPTTPKEPAPTTTKEPAPTTPKEPAPTTPKKRAPTTPKEPAPTTPKEPA 308
            :||||                     |||                     :|||
Db  1364 TSTSSTTTP--PPTTTDLPPISTTGLP-PTTTELPPT------ -TTTDLPPTTTRLPP 1413

Qy   309 PTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTPK------ -EPSPTTTPKEPAP    363
            :|||                |||                    :||
Db  1414 TTTTSLPPTTTTGLPPTTTTGAQPTTTTLSSEFETSVTTTSPESTTQPPSTTTMKPLPAG 1473

Qy   364 ------ -TTPK------ -EPAP--TT                                373
                     :||         :||
Db  1474 TECTGEGYMADPEDCRKYYRCINAGASYRKYNFTCPKGTGWNEEVCGTCDYVENIPRCSKL 1533

Qy   374 ------ -KEPAPTTPKE-PAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPT       428
                   :||||       |||
Db  1534 PKKPAPTTP------                                             1591

Qy   429 TPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKE  488
            :||                                               |||
Db  1592 TTEE--PEKPQKPT-TTEYPQKPTTTEEPTPEKPQKPT-TTEYPQKPTTTEEPTPEKPQK 1647

Qy   489 PAPTT-PKEPAPTTPKETA-------PTTPK--GTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
            :||||                     |||
Db  1648 PTTTEYPQKPTTEEPTTTSIPGYNPTTTSVPGYNPTTTPIPVETTTSTPGYK---PTTTG 1704

Qy   541 EPTSTTPTTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKK 597
            :||                     |||                     :|||
Db  1705 EPITTTNT-LPSTTTDAIQEPTTSKKPEPTTTTESPESSTPEGSVTTLQPEPQPNYNCSSE 1763

Qy   598 ---PAPKELA------ -PTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTP 648
               :||                     |||
Db  1764 GFFPDEDCSRYYRCVDAARNGKYQVYAFKCGKGTVWDTSTETCNYADQVSGN------ 1816

Qy   649 ETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPCGVPTTKTPAAT 708
            :||                                               |||
Db  1817 --CSSGQTTTPGTTTTEPGTTESTTSSGKP------ -ETTSKAPEN-----        1857

Qy   709 KPEMTTAKDKTTERDLRTTPETTTPAAPKMTKEPATTTEKTTTESKITATTTQVTSTTTQD 768
            :||                  |||                  :||
Db  1858 -PETTTT------ -SSPETTTT------ -VASETTTTSGT------ -TTTATPETTTKP 1895

Qy   769 TTPKRITTLKTTTLAPKVTTTLKITTTTTEIMNKPEETAKPKDRATNSKRATTPKPQKPTKA 828
            :||                                               |||
Db  1896 PKP------ -ETTTIAGEETSTSKSPTTTE------ -SPAPSTNTSA              1928

Qy   829 PKKPFSTKKPKTMP 842
```

```
                           |
                           |:||
Db   1929 p-------CPETGP 1935
```

RESULT 14

Q9XDH2

```
ID   Q9XDH2;           PRELIMINARY;      PRT;    763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH MUCIN HOMOLOG.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang Y., Moreno C., Singh M.;
RT   "Cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003882; Pistil_extensin.
DR   PRINTS; PR01222; ATROPHIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE   763 AA;  39168EC45A5916F8 CRC64;

Query Match      13.9%;  Score 715;  DB 2;  Length 763;
Best Local Similarity  28.2%;  Pred. No. 2.7e-38;
Matches  248;  Conservative  39;  Mismatches 361;  Indels 232;  Gaps 39;

Qy   132  PTPKAETTPKGPALTTPKEPTPTPTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTKSAP 191
               :|||    ::|                        :||
Db   3    PVP-----APRALAPLPPAPPAPAEPKSKPFPPAP----------PAPPCMMLVSAAP 46

Qy   192  TTPKEPAPTTTKEPAPTTPKEPAPTTKSAPTTPKEPAPTTPKEPAPTTPKE 251
               |||
Db   47   PCP-PAP----PAPPKPKSKAPPPVPPAPPARELAPPLP--PAP-----PEAPRE 90

Qy   252  PAPTTPKEPTPTT----PKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK 305
Db   91   SRPALPPCPEPPVVIPDPPEPAAPPVPAPNSPPFPPPPPAPKPVPAAPVP--PVPNSPP 148

Qy   306  EP--APTTTKEPSPTTPKEPAPTTKSAPTTPKEPAPTTPKEPSPTTPKEPAP 363
Db   149  FPPPPAALNPPAP-----PAPPLANSPPLPPPAPPTPAGT--PPAAPWPVPAAPKSKPA 201

Qy   364  TTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAP----TTTPKKPAPTTAPKEPAPTTPKETAP 419
Db   202  SPPRPPAP-----PMPATPMEFPPLPPVPPDPISKETPPAPPAPPIPAPVPIPPVPPLP 256

Qy   420  TTPKKLTPTTPEKLAP------TTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTT-- 469
Db   257  PVPNKIPPAPP---APPVAVAVLVAPCPLPLPNNHPAPPAAPVPGVPLAPLPNSHP 313

Qy   470  ---PKAAANTPKEPAP-------------PKEPAPTTPKEPAP 483
Db   314  PAPPSAPVPGVPLAPLPISGRPVSVWKGSFTTLSTFCCRVCSGEVLAGALNPSRPSRSPL 373

Qy   484  --TTPKEPAPTTPKKPAPTTPKEPAPTTPKETA----PTTPKGTA---PTTLKEPAPTTP-KKPAPKE 533
Db   374  TTTTPALPAPIPPLPPLPLPINTAVPPIPPLPVTALAPPLPPLAPLPISPGVPPAP-- 431

Qy   534  LAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPA----PTTPKEPAPTTPKGTAPTTL 587
Db   432  --PIPPGKPWTTPPLAPAPPEPK-TVPVLPGPSCPPSEKPNPAPPEPEPPEPKSSPALPP 488
```

RESULT 15

Q9FPQ6

```
ID   Q9FPQ6;           PRELIMINARY;      PRT;    555 AA.
AC   Q9FPQ6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VEGETATIVE CELL WALL PROTEIN GP1.
GN   GP1.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA   Goodenough U.W.;
RT   "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT   plant hydroxyproline-rich glycoproteins.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
EMBL; AF309494; AAG45420.1; -.
DR   InterPro; IPR003882; Pistil_extensin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE   555 AA;  6A584A90465502F5 CRC64;

Query Match      13.7%;  Score 707.5;  DB 10;  Length 555;
Best Local Similarity  28.9%;  Pred. No. 6.2e-38;
Matches  189;  Conservative  66;  Mismatches 243;  Indels 157;  Gaps 22;

Qy   588  KEPAPTTP---KKPAPKELAPTTTKGP-TSTTSDKPAPTTPKET--APTTPKEPAPTTP 640
          |||:|                :|       |||:                    ||
Db   489  APPAPSMPSAVRVPPSPPIPPAPPAAPRASMPALPPAPPSPPATRLCPLPPSPPAPNSP 548

Qy   641  KKPAPTTPETPPPTTPSEVSTPTTTPKEPTTIHKSP----DESTPELSAEPTPKA--LENSP 694
Db   549  -PAPPAPPIPPKLLS--ANPCPPVPPAPNRPPAPPAPPELPAPPDPPTPPVANSP 604

Qy   695  KEPGVPTTPKIPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKI 754
Db   605  --PAPPAPPAPSALPFVNPPA----------PPTPAAPK----SRPAL 637

Qy   755  TATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATN 814
Db   638  PAAPPAPPAPPVRATTP-------------PPAPPAPPAPN 665

Qy   815  SKATTTPKPQKP----TKAPKKPTSTIKKPKTIMPRVRKPKITTPTRKMTSTIM----PELNP 865
Db   666  SMALPPAPPDPPIPLLATPPAPPLPMSPPAPPLPPAAPDPPAPPLITINQPPSPPLAP 725

Qy   866  TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET 905
Db   726  VPGAPLAPLPINGRPVFARKNSLI----------GSSSGDT 756
```

```
Qy   434  APTTPEKPAPTTPEEPAPTTPEEPTPTTPKAAAPNTPKEPAPTTPKEPAPTT
          ||  |  |||  :||  |  |||:|  |  ||| ::|  ||  |:|     |:|
Db   195  APPVPPSPAPPSP---APPVPPSPAPPSPSPSPAPPSPSPSPAPPSPVPPSP-A  250

Qy   494  PKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTIKEPTSTTSDKPAPT  553
          |  |||  |  ||  :|| ::|  ||  | |:
Db   251  PPSPAPPSPKPPAP-------PPPPSPPPPPRPPPFPANTPMPPPSPSPPPPSPA  298

Qy   554  TPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTS  613
          |            |:|        |||        |:|
Db   299  PPTPPTPPSPSPPSP-VPPSPAPVPPSPAPPS----PAPSPPPSPAP----PTPSPSPSP  349

Qy   614  TTSDKPAPTTPKETAPTTPKKPAPTTPETPPPPTTSEVSTPTTTKEPTTIHKS  673
          |:|:|       |:|:|       :|:|:|       |:|:|
Db   350  SPSPSPSPS----------PSPSPSPSPSPIPSPSPKPSPSPVAVKLVWADDAI----  393

Qy   674  PDESTPELSAEPT-PKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETT  732
          :  |:   |  | | |
Db   394  ---AFDDLNGTSTRPGSASRMVGEPDIAGTKCKGNLKGWMPKPSRNPRWGQAVFSCGGRTV  450

Qy   733  TAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTK--  790
          |||  |||  | ::  |       |  |:|
Db   451  GSVANVTIRVAPATEK--PALIYSSIELVVYNTG--------ATLIRVPIAANVTRSQIR  500

Qy   791  ---KITTTTEIMNKPEETAKPKDRATNSKATT-PKPQKPTKAPKKPTSTKKPKT  840
                |||||
Db   501  CPGFLTYGTTPIAGYP-----------TGIDATTWPNWKIAGVRINMGAGNKKPKT  545
```

```
                        GenCore version 4.5
                        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:                 April 26, 2002, 16:13:35 ; Search time 153.5 Seconds
                        896.692 Million cell updates/sec

Title:                  US-09-556-246-1_COPY_200_1140
Perfect score:          5011
Sequence:               1 VKDNKKNRTKKKPPTPKPPVV...........DMDYLPRVPNQGIIINPMLS 941

Scoring table:          BLOSUM62
                        Gapop 10.0 , Gapext 0.5

Searched:               473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters:    473505

Post-processing:        Minimum Match 0%
                        Maximum Match 100%
                        Listing first 45 summaries

Database :              SPTREMBL_17:*
                        1:  sp_archea:*
                        2:  sp_bacteria:*
                        3:  sp_fungi:*
                        4:  sp_human:*
                        5:  sp_invertebrate:*
                        6:  sp_mammal:*
                        7:  sp_mhc:*
                        8:  sp_organelle:*
                        9:  sp_phage:*
                        10: sp_plant:*
                        11: sp_rodent:*
                        12: sp_virus:*
                        13: sp_vertebrate:*
                        14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                    SUMMARIES
                  %
Result          Query
No.   Score   Match  Length  DB  ID          Description
----  -----   -----  ------  --  --          -----------
  1   5011   100.0    1404    4  Q92954      Q92954 homo sapien
  2   5002    99.8    1404    4  Q9BX49      Q9BX49 homo sapien
  3   1904    38.0    1054   11  Q9Jm99      Q9Jm99 mus musculu
  4    967    19.3    1079   11  Q9N4S7      Q9N4S7 caenorhabdi
  5    936.5   18.7   1049    5  Q917S1      Q917S1 drosophila
  6    865.5   17.3   1795    5  O76894      O76894 drosophila
  7    853     17.0   1489   10  Q96449      Q96449 phytophthor
  8    806.5   16.1   1315   10  Q9SPM0      Q9SPM0 zea mays (m
  9    802     16.0   1274    5  Q2O007      Q2O007 caenorhabdi
 10    800.5   16.0   1480   10  Q9LIE8      Q9LIE8 arabidopsis
 11    784     15.6   1188   10  Q41805      Q41805 zea mays (m
 12    776     15.5   2187   11  P70670      P70670 mus musculu
 13    739     14.7   2112    5  Q9VEL9      Q9VEL9 drosophila
 14    715     14.3    763    2  Q9XDH2      Q9XDH2 mycobacteri
 15    707.5   14.1    555   10  Q9FPQ6      Q9FPQ6 chlamydomon
 16    671     13.4   1151   13  O57580      O57580 gallus gall
 17    667.5   13.3   4880   11  Q9JLT1      Q9JLT1 rattus norv
 18    667.5   13.3   5085   11  Q9JKS6      Q9JKS6 rattus norv
 19    660.5   13.2   5677    5  Q9N435      Q9N435 caenorhabdi
```

```
20   659.5   13.2   1229    5  Q94185      Q94185 caenorhabdi
21   654     13.1   2284    5  Q9VPG1      Q9VPG1 drosophila
22   653.5   13.0   1514    5  Q9GUW7      Q9GUW7 leishmania
23   647     12.9   4833   11  Q9QYX6      Q9QYX6 mus musculu
24   647     12.9   5038   11  Q9QYX7      Q9QYX7 mus musculu
25   641.5   12.8   2089    4  O14676      O14676 homo sapien
26   633     12.6    489   10  Q41707      Q41707 homo sapien
27   632     12.6    762   10  Q92OIO      Q92OIO arabidopsis
28   632     12.6    7962    4  Q104465     Q104465 homo sapien
29   629     12.6   3507   10  Q104465     Q104465 homo sapien
30   629     12.6   6632    5  Q17362      Q17362 caenorhabdi
31   626.5   12.5    990   13  Q91803      Q91803 xenopus lae
32   624.5   12.5   1612    5  Q9VQO2      Q9VQO2 drosophila
33   624.5   12.5   2768    5  Q9VCO0      Q9VCO0 drosophila
34   622.5   12.4    839    2  Q9RX57      Q9RX57 deinococcus
35   622     12.4   6642    5  O01761      O01761 caenorhabdi
36   620.5   12.4    971    5  Q9XVS4      Q9XVS4 caenorhabdi
37   617     12.3   3570    4  Q99552      Q99552 homo sapien
38   607.5   12.1    801   10  O23635      O23635 caenorhabdi
39   607.5   12.1    924   12  Q99307      Q99307 epstein-bar
40   606.5   12.1   2344    5  Q9N3Y8      Q9N3Y8 caenorhabdi
41   605     12.1    379    5  O27929      O27929 drosophila
42   600.5   12.0   1893    5  Q9NKC9      Q9NKC9 drosophila
43   598.5   11.9    409   10  Q9SBM1      Q9SBM1 volvox cart
44   577     11.5    956   10  Q9LJ64      Q9LJ64 arabidopsis
45   573     11.4   2601    4  Q9BZ84      Q9BZ84 homo sapien
```

```
RESULT  1                                              ALIGNMENTS

ID   Q92954             PRELIMINARY;     PRT;    1404 AA.
AC   Q92954;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEGAKARYOCYTE STIMULATING FACTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Blood 78:279-279(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA   FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA   Jacobs K., Turner K.;
RL   (in) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL   Mosher D.F. (eds.);
RL   Biology of vitronectins and their receptors., pp.45-52,
RL   Elsevier Science Publishers B.V. (1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RL   EMBL; U70136; AAB09089.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   InterPro; IPR002400; GF_cysknot.
DR   Pfam; PF00045; hemopexin; 2.
DR   Pfam; PF00086; Somatomedin_B; 2.
DR   Pfam; PF01033; GFCYSKNOT.
DR   PRINTS; PR00438; GFCYSKNOT.
```

```
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS005524; SOMATOMEDIN_B; 2.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
SQ   SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match             100.0%;  Score 5011;  DB 4;  Length 1404;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches  941;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy        1 VKDNKKNRTKKKPKTPKPPVVDEAGSGLDNGDFKVTTPDTSTSTTQHNKVSTSPKITTAKPIN 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      200 VKDNKKNRTKKKPKTPKPPVVDEAGSGLDNGDFKVTTPDTSTSTTQHNKVSTSPKITTAKPIN 259

Qy       61 PRPSLPPNSDTSKETSLIVNKETTVEKTKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      260 PRPSLPPNSDTSKETSLIVNKETTVEKTKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 319

Qy      121 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPPTTIKSAPTTPK 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      320 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPPTTIKSAPTTPK 379

Qy      181 EPAPTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKSAPTTPKERAPTT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      380 EPAPTTIKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKSAPTTPKEPAPTT 439

Qy      241 PKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEPA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      440 PKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEPA 499

Qy      301 PTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPSPTTTKE 360
            |||||||||||||||||||||||||||||||||||||||||||
Db      500 PTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPSPTTTKE 559

Qy      361 PAPTTPKEPAPTTPKKPAPTTPKKPAPTTPKEPAPTTPKETAPT 420
            ||||||||||||||||||||||||||||||||||||||||||
Db      560 PAPTTPKEPAPTTPKKPAPTAPKEPAPTTPKEPAPTTPKETAPT 619

Qy      421 TPKKLTPTTPEKLAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 480
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      620 TPKKLTPTTPEKLAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 679

Qy      481 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      680 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 739

Qy      541 EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKKPAPTTPKKPAP 600
            ||||||||||||||||||||||||||||||||||||||||||
Db      740 EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKKPAPTTPKKPAP 799

Qy      601 KELAPTTPKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTISEVST 660
            |||||||||||||||||||||||||||||||||||||||||||
Db      800 KELAPTTPKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTISEVST 859

Qy      661 PTTTKEPTTIHKSPDESTPELSAEPTTKPALENSPKEPGVPTTKIPAATKPEMTTTAKDKT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      860 PTTTKEPTTIHKSPDESTPELSAEPTTKALENSPKEPGVPTTKIPAATKPEMTTTAKDKT 919

Qy      721 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTISTTTQDTTPFKITTLKTT 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      920 TERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTISTTTQDTTPFKITTLKTT 979

Qy      781 TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKT 840
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      980 TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKT 1039

Qy      841 MPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG 900
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1040 MPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG 1099

Qy      901 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 941
            |||||||||||||||||||||||||||||||||||||||
Db     1100 AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1140
```

```
RESULT 2

ID   Q9BX49
AC   Q9BX49;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).
GN   BG174L6.2
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133553; CAC36090.1; -.
DR   EMBL; AL174L6.2
SQ   SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;

                        PRELIMINARY;   PRT;   1404 AA.

Query Match              99.8%;  Score 5002;  DB 4;  Length 1404;
Best Local Similarity    99.8%;  Pred. No. 0;
Matches  939;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy        1 VKDNKKNRTKKKPKTPKPPVVDEAGSGLDNGDFKVTTPDTSTSTTQHNKVSTSPKITTAKPIN 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      200 VKDNKKNRTKKKPKTPKPPVVDEAGSGLDNGDFKVTTPDTSTSTTQHNKVSTSPKITTAKPIN 259

Qy       61 PRPSLPPNSDTSKETSLIVNKETTVEKTKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      260 PRPSLPPNSDTSKETSLIVNKETTVEKTKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAK 319

Qy      121 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPPTTIKSAPTTPK 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      320 DLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPPTTIKSAPTTPK 379

Qy      181 EPAPTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKSAPTTPKEPAPTT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      380 EPAPTTIKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKSAPTTPKEPAPTT 439

Qy      241 PKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEPA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      440 PKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEPA 499

Qy      301 PTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPSPTTTKE 360
            |||||||||||||||||||||||||||||||||||||||||||
Db      500 PTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPSPTTTKE 559

Qy      361 PAPTTPKEPAPTTPKKPAPTTPKKPAPTTPKEPAPTTPKETAPT 420
            ||||||||||||||||||||||||||||||||||||||||||
Db      560 PAPTTPKEPAPTTPKKPAPTAPKEPAPTTPKEPAPTTPKETAPT 619

Qy      421 TPKKLTPTTPEKLAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 480
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      620 TPKKLTPTTPEKLAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKE 679

Qy      481 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db      680 PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTK 739

Qy      541 EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKKPAPTTPKKPAP 600
            ||||||||||||||||||||||||||||||||||||||||||
Db      740 EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKKPAPTTPKKPAP 799

Qy      601 KELAPTTPKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTISEVST 660
            |||||||||||||||||||||||||||||||||||||||||||
Db      800 KELAPTTPKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTISEVST 859
```

```
Qy   661  PPTTKEPTTIHKSPDESTPELSAEPTPKALENSKPEPGVPTTKIPAAATKPEMTTTAKDKT  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   860  PTTTKEPTTIHKSPDESTPELSAEPTPKALENSKPEPGVPTTKIPAAATKPEMTTTAKDKT  919

Qy   721  TERDLRTTPETTTTNAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   920  TERDLRTTPETTTTNAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT 979

Qy   781  TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKAATTPKPQKPTKAPKKPTSTKKPKT  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   980  TLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKAATTPKPQKPTKAPKKPTSTKKPKT  1039

Qy   841  MPRVRKPKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1040  MPRVRKPKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGG 1099

Qy   901  AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS  941
          |||||||||||||||||||||||||||||||||||||||||
Db  1100  AEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS  1140
```

RESULT 3

Q9JM99   PRELIMINARY; PRT; 1054 AA.

```
ID   Q9JM99
AC   Q9JM99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE   PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN   PRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ikegawa S., Nakamura Y.;
RT   "a novel mouse gene gene highly homologous to a human gene encoding
RT   megakaryocyte stimulating factor precursor and cartilage superficial
RT   zone protein.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034730; BAA92310.1; -.
DR   MGD; MGI:1891344; Prg4.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ   SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;
```

Query Match   38.0%;   Score 1904;   DB 11;   Length 1054;
Best Local Similarity   44.4%;   Pred. No. 3.5e-113;
Matches   423;   Conservative 33;   Mismatches 128;   Indels 368;   Gaps 20;

```
Qy    1  VKDNKKNRTRKKKPPPKPPVVDEAGSGLDNGDFKVT--TPDTSTTQHRNKVSTSPKITTAKP  58
Db  195  VKDNKKNTPKKKPNPEPPAVDEAGSGLDNGEFKLTPPPDPDPTTPHSKVATSPKTTAAKP  254

Qy   59  INPRPSLPPNSDTSKETSLTVNKETTTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTS  118
Db  255  VTPKPSLAPNSETSKEASLASNKETTVETKETTATNKQSSA-SKKKTTSVKETRSAEKTS   313

Qy  119  AKDLAPTSKVLAKPTPKAETTKGPALTTPKEPTTPTPKEPASTTPKEPTPTTIKSAPTT    178
Db  314  DKDV-------------------------EPTSTTPK--------NSAPTT          331

Qy  179  PKEPAPTTPKKPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAP 238
Db  332  TKKPV-TTTKESKFLP-----------LPQEPEPTTAKPPPTTTKKPEPTTKEPEP       376

Qy  239  TTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE  298
Db  377  TTPKEPEPTTPKEPEPTTPKEPAPTTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTT  436

Qy  299  PAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKSAPTTTKEPAPTTTKSAPTTTKEPSPTTT   358
Db  437  PEPTTTKEPESTTTRKEPEPTTPKEP--------------EPTTPKEPEPTTL         474

Qy  359  KEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTKKPAPTAPKEPAPTTPKETA   418
Db  475  KEPEPTTPKEPEPTTPKEPEPTTPKEPEPTTPKEP---------------           512

Qy  419  PTTPKLTPTTPKEKLAPTTPKPKPAPTTPEELAPTTPEEPTPTTPKEPEPAPTTPKAAAPNTP 478
Db  513  ---------------------------------------------------TP          514

Qy  479  KEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPAPKELAPTT    538
Db  515  KEPEPTTPKEPEPTTPKEPEPTTP----------KEPEPTTPKKP-----           549

Qy  539  TKEPTSTTSDSKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKP  598
Db  550  ------------EPTTPKEPVPTTPKEPEPTTP------KEPEPTTPKEP           581

Qy  599  APKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKRPAPTTPETPPPTTSEV   658
Db  582  EP--------TTPKEPTTPKEPTTPKEPTTPKEPTTP-----------           617

Qy  659  STPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAATKPEMTTTAKD  718
Db  618  -----------------------                                  617

Qy  719  KTTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLK  778
Db  618  ---------TSPKT---------                                  626

Qy  779  TTTLAPKVTTTKKITTTEIMNKPEETAKPKDRATNSKAT-------------TPKP-QKPTAKP 829
Db  627  AYTTLAPKVTAPAE-----EIQNKPEETTPASEDSDDSKTTLKPQKPTKAPKPTKKPTKAP  681

Qy  830  KKPISTKKPKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLV 889
Db  682  KKPISTKKPKT-PKTRKPKTTPAPLKTISATPELNTTP--LEVWLPTTIPKQTPNPETA   738

Qy  890  EVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS          941
Db  739  EVNPDHEDADGEGEGEKP-LIPGPPVLFPTAIPGTDLLAGRLNGININPMPS          789
```

RESULT 4

Q9N4S7   PRELIMINARY; PRT; 1079 AA.

```
ID   Q9N4S7
AC   Q9N4S7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y51B11A.1 PROTEIN.
GN   Y51B11A.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
```

```
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Cotton M.;
RT   "The sequence of C. elegans cosmid Y51B11A.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE  1079 AA;  110532 MW;  8DBDE3824CF80CA1 CRC64;
```

```
Query Match         19.3%;  Score 967;  DB 5;  Length 1079;
Best Local Similarity  32.3%;  Pred. No. 8.4e-54;
Matches  294;  Conservative 114;  Mismatches 377;  Indels 124;  Gaps 33;
```

```
Qy   35 TTPDTSTTQHNKVSTSP-KITTAKPINPRPSLPPNSDIS--KETSLTVNKETIVETKETT  91
         |||::|| :|| :||  :|||  |  :||   ::||:||:      :||| |:| |::||
Db  102 TAPETTSTEAPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEAPSSS 161

Qy   92 TTNKQTPTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGP--ALTTPK 149
         || |:| :| |  :| : | |||   |  ||| | : ||| | |  :|:||    |||||
Db  162 TTPVQTT-----------TTTAPETTSTEPPSSS---TSPVQTTTTTAPETTSTEPPSSSTTPV 211

Qy  150 EPTPTTPKEPASTTPKEPPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTT 209
         |:||||||||:  :||||  |||:|||||||||||||||||||||||||||||:||||||
Db  212 QTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPE----TT 268

Qy  210 PKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTTKKPAPTTTKEPAPTTTKEPTPTTPKEPA 269
         ::||||||||||||||:|:|||||||||||||  ||||||||||||-|:|:  :||||
Db  269 STEP-PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEP-PSSSTTPV 326

Qy  270 PTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTTKEP-------------APTTTKEP 315
         ||||||||:||||  |  |:||||| | ||||||||||             |||||||
Db  327 QTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTPVQNTTTTAPETTSTE 386

Qy  316 SPTTPKEPAPTTTKSAP---TTTKEPAPTTTKSAPTTPKEPSTTTPKEP-APTTPKEPAPTT 372
         |:|||||||||:|:|     :|||||||||||  :|||||  :||| :||||||||:|:
Db  387 PPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTT 446

Qy  373 TPKKPAPTTPKEPAPTTPKEPAPTTAPKEP--APTTPKE----TAP------- 419
         |||||||||||||  ||:|||    ||:|    ||||||    |||
Db  447 T----APETTSTEPPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEP 502

Qy  420 ----TTPKKLTPTTPEKLAPTTPEKPAP---TTPEELAPTTPEEPPTTPEE------PAPT 468
             :|||:|||||| |||  ||      :||:|| |||:|    ||:||      ||||
Db  503 PSSSSTTPVQTTTTT----APETTSTESPSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQT 558

Qy  469 TPKAAANTPKEP---APTTPKEPAPTTPKEPAPTTPKETAPTTLKEPAPTTP 526
         |:|  | |:|||   :|||||||||||||  |||||||  ||| |||||||
Db  559 TTTTTAPETTSTEPPSSSTTPVQTTTTTTAPETTSTEPP--APTTPKGTA 583

Qy  527 KKPAPKELAPTTKEPTSTTSDKP-APTTPKETAPTTTPKEP---APTTPKGTA 583
         |||:||| ||| |: :||:||   :|||||||||||||||    ||||||:||
Db  584 PTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKP-APTTPKEPAPTTTPKK 642

Qy  618 PSSSTTPVQTTTTTTAPETTSTEPPSSSTTPVTAPE---TTSTEPPSSSTTPVQTT 674
         |:: ||||||||||||||||||||||||||:||||    :|||||||||||||||
Db  643 PAPTTPEIPPPTTSEVSTPTTTHKSPDESTPELSAEPT---PKALENSPKEPGV 699

Qy  643 PAPTTPEIPPPTTSEVSTPTTTHKSPDESTPELSAEPT---PKALENSPKEPGV 699
         |:: ||||||| :|:::|||| | :  ||  |||:|    |||    ||  :|
Db  675 TTTTAPETTSTEPPSSSNTPVQTTTTTTAPETTSTEPP-TTSTEPPSSSTTPVQTTTT 732

Qy  675 TTTTAPETTSTEPPSSSNTPVQTTTTTTAPETTSTEPP-TTSTEPPSSSTTPVQTTTT 732
         ||  |:| | |:| | :||  |||||:||| || |:|     ||:|| |||||:||||
Db  733 --PSSSTTPVQTTTTTTAPETTSTEPPSSSNTPVQTTTTTTAPETTSTEPPSSSNTPVQTT 786

Qy  700 PTIKTPAAIKPEMTTTAKDKTTTERDLRTT----PETTTAAPKMTKETATTTEKTTESKIT 755
         |:| :|||: |:|:|||:|:||:::|:|:|    |||| :||| :| | |:|: |:|:|
Db  787 QTTTT---TAPETTSTEPPSSSPVQTTTTTTAPETTSTEPPSSSNTPVQTTTTTAPETT 843

Qy  733 --PSSSTTPVQTTTTTTAPETTSTEPPSSSNTPVQTTTTTTAPETTSTEPPSSSNTPVQTT 786
```

```
RESULT  5
Q9I7S1
ID   Q9I7S1             PRELIMINARY;      PRT;   1049 AA.
AC   Q9I7S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003495; AAG22253.1; -.
DR   Flybase; FBgn0030561; CG5228.
SQ   SEQUENCE  1049 AA;  107278 MW;  954DD629E7619671 CRC64;
```

```
Qy  756 ATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTK-----KTITTTEIMNKPEETA-KPK  809
         :|:||:    :||: ::||:     |||||||:||     |:|||:::|||||| ||||
Db  844 STEPPSSSTTPVQT---------TTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPP  894

Qy  810 DRATNSKATTPKPQKPTKAPKKPTSTKKPKIMPRVRKPKTTPTPRKMTSTMPELNPTSRI  869
         : |                 :|:||| ||:||||  |||||||||||||||||:::|:
Db  895 SSSTTPVQTTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTP--------  946

Qy  870 AEAMLQTTT  878
         :||:||||
Db  947 ----VQTTT  951
```

Query Match   18.7%;   Score 936.5;   DB 5;   Length 1049;
Best Local Similarity   34.6%;   Pred. No. 7e-52;
Matches 371;   Conservative 45;   Mismatches 439;   Indels 217;   Gaps 54;

```
Qy      2 KDNKKNRTKKKPPTKPPPVVDEAGSGLDNGDFKVTTPDTS---TTQHNKVSTSPKIITTAKP 58
            ||| | |||                            |||  ::
Db     30 KPNEIGTTAKPTTLKP--------------------------TEGTSAKPTTLKPTEGTSAKPTTLKP 71

Qy     59 -----INPRRSLGPNSDTSKETSL--------TVNKETTVETKETTTTNKQT--STDGKEKTT 106
                |:||       |    | |:|:   |  ||| ::
Db     72 TEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG----T 127

Qy    107 SAKET--QSIEKTSAK---DLAPISKVLAKPT-------PKAETTKGPALTTPKEPTPTT 155
            ||| |               |                | |  |||
Db    128 SAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG 187

Qy    156 PKE-----PASTTPKEPT---PTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAP 207
            |||        |  |  |     |::|||||
Db    188 PTEGTSAKPTTLKPTEGTSAKPTTLK--PTEGTSAKPTTLKPTEG 243

Qy    208 TTPKEPAPTTTKEPAPTTTKSAPTTPKE--PAPTTTPKKPAPTTTPKEPAPTTTPKEPT---P 262
            || ||||||||||||| | :||||||||     |||  ||||||||    :
Db    244 TTAK---PTTLKPTEGTTAKPTTLNPTEGTSAKPTTLKPTEGTSAKP 300

Qy    263 TTPKEPAPTTTKEPAPTTTPKEPAPTAKPKKPAPTTTPKEPAPTTTPK-----EPAPTTTKEPSP 317
            || ||||||||| ||             ||||||||    |
Db    301 TTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTDG 360

Qy    318 TTPKEPAPTT---------TKSAPTTKEPAPTTTTKSA---PTTPKEPSPTTTKEPAPTTPKE 368
            || |||||||          ||:||||||||           || ||| |||||||||||
Db    361 TTAK---PTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTTAK- 416

Qy    369 PAPTTPKKPAPTTTPKEPAPTTTPKEPAPTTTKKPA---PTAPKEPAPTTPKETAPTTTPKGTA------ 423
            ||||| |||:        ||||   ||   |    ||           ||:|||||||||
Db    417 --PTTLKPTEGTTAK----PTTLK-PTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTTA 470

Qy    424 KLTPT------------TPEKLAPTTPEKPAPTTPEELAPT---TPEEPTTTPEEPAPTTKAAAPN 476
            |:|||             || |  ||| ||||          |||      |   ||  ||
Db    471 TLKPTEGTTAKPTTLKPTKGTSAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPT 530

Qy    477 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTTPKGTA--------------PTTLKPEPAPTTP 526
            |||||||||||||||||| |||||||| |              |||||||
Db    531 KGTTAKPTTLK---PTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTTA 587

Qy    527 K-----KPA------------TTKEPTSTTSDKPA----------------PTTPKGT 558
            |       |||           ||:||||   |||                  |||
Db    588 KPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPT 647

Qy    559 APTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPK---KPAPKELAPTTPKGPTST 614
            || | ||| ||||  |  ||||           |        || |:|| |||| |||
Db    648 EGTTAK---PTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTEG 701

Qy    615 TSDKPAPTTPKE--TAPTTPKEPAPTTPKKPA--PTTPETPPPTSEVSTPT-TTKEPT 668
            |||| ||||||     |||||   ||||||     || |::|||||    :|||||
Db    702 TSAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGRTTAKPT 761

Qy    669 TIHKSPDESTPELSAEPTP--------------KALENSPKEPGV--PTTKT---PAATK 709
            |:|      ||  | ||               ||: | ||     |||||     |||
Db    762 TL--KPTEGT----SAKPTTLKPTEGTTAKPTTLKPTERTSAQPT 816

Qy    710 PEMTTTAKDKTTERDLRTT----------------PETTTAARKMKETATT----------TEKTTES 752
            ||||||||||||        |              || || ||  |:|           |:|||:
Db    817 PTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTTAKPTTLKPTEGTSAKP 876

Qy    753 KITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKK---TITTTEIMNKPEETAKPKD 810
            |:| ||| |: |||   ||  |||:|||| |  |  |||   :||| | ||| | ||
Db    877 PTTLKPTEGTSAKPTTLKPFDGTTAKPTTLKPTEGTSAKPTTLKPTE 936

Qy    811 RATNSKATTPKPQKPTKAPK---KPT--STKKPKTMPRVRKPKTTPT---PRKMTTSTMP- 861
            |:|||:| |||  || || |   |||  |:||| |       ||||       :|:|:|
Db    937 -GTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGPSAKPTTLKPTERTSAQPT 995
```

RESULT 6

```
DR   FlyBase; FBgn0025390; EG:5G7.1.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR002965; P_rich_extensn.
Pfam;  PF01607; Chitin_bind_2; 2.
PRINTS; PR01217; RICHEXTENSN.
SMART; SM00494; ChtBD2; 2.
DR   SMART; SM00494; ChtBD2; 2.
SQ   SEQUENCE  1795 AA;  194464 MW;  07F10C129BD9557B CRC64;

Query Match          17.3%;  Score 865.5;  DB 5;  Length 1795;
Best Local Similarity  28.5%;  Pred. No. 3.5e-47;
Matches 276;  Conservative 129;  Mismatches 447;  Indels 115;  Gaps 28;

Qy     6 KNRIKKKPPPKPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL 65
           ||| |||
Db   431 KNDVTEAPEIKSPL----------KGLIHLSENIVILPETTTT--TTTTKPVVLTCPTISPPDTT 483

Qy    66 PPNSDT----------------SKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSI 114
          |  | |                |  || || | |||  ||  |  || | |||
Db   484 PKPSTTTAVTKSTPKISSTEQHSTTTAKTTTTKRPTTVVEKTSSATEKPRTTVVTTTTQK 543

Qy   115 EKTSAKDLAPTSKVLAKPPTPKAETTTKGPALTTTKKEPTPTTPKEPASTTPKEPATTTKS 174
          |:||  ||||  |  ||  | ||  :   ||: :| |
Db   544 RSTTHNTSPDTKTTIRSTTLSPKVTTTPSTTPTTPST-- 594

Qy   175 APTTPKEPAPTTTKSAPTTPKEPAPTTTK-EPAPTTPKEPAPTTTKSAPTTP 233
          :|  |||  | | ||    | ||  |||    | |||  | |  | |  |
Db   595 --TTPSTTTTPSTTTTVKVSTTHRPRTTSQKTTTASTTTKK----TTTSPKTTKTTDIPTST 648

Qy   234 KEPAPTTPKRPAPTTPKEPAPTTPKEPAPTTTTPKEPAPTTPKEPAPTTPKEPAPTAPKKPAP 293
          |||||:  | | |||  | |||  | |||  | |
Db   649 TSKLSTTTTQKTTTTTHKFTAATTSTEKPKTTTTEKTSTVSTTVSTTTKKSTESSSPKPTSSTGKPT 708

Qy   294 TTPKEPAPTTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 353
          ||  |||| | || |||  | |  |||  |  |||  |  |  |||
Db   709 TTPKPSTRTTPPTTKVTTTQITTTTPLRSSTETTTSTQPPTTTTPQPTTTTTLTVTPKTS 768

Qy   354 SPTTTKEPAPTTTPK------KPAPTTPKEPAP 392
          || ||| |||   |      ||| ||||
Db   769 TTTTTTEKPTISSPKPTTTTQKTTSTTAPNTTKVAITTQKETTPTQSTSTTTFTRKTTTTNN 828

Qy   393 RAPTTTKKPAPTTAPKEPAPTTPK-ETAPTTTPKKLTPTTPEKLAPTTPEKP 443
          | | | | || |    | ||  |    |||    |  || | |  | |
Db   829 PEPTSTEKPITSTTTPKPSTTPKTSTVASSTEKTTISSPKPTTEKSTENPTTNSVKTTSAL 888

Qy   444 TTPEELAPTTPEEPTTPEEPAPTTPKPKAAAPNTPKPKEPAPTTPKEPAPT--TPKEPAPTT 501
          ||  |||  |    | | ||| ||   | ||  |  | |||  ||| |||
Db   889 TSSTQRATSTTSEPTKTT-QNITTTPKPTTLKTSTQEATTSTQKVSTVTTTKKATESS 947

Qy   502 PKETAPTTPKGTAPTTLKEPAPTTPKPAPKELAPTTTKEP:--STTSDKPAPTTPKGTA 559
          ||| |||| | |||  |||||| |  |       | :| |:  ||: | ||| ||||
Db   948 PLTTLSTEEPNTPKPLRTTTPTTTSVATTATTRITTTTTISESSTETTSTQKPKSTTTPTSTT 1007

Qy   560 PTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP-TTTKGPTSTTSDK 618
          || |||| | ||| |||  |  | |||  |||  ||  |   |    |||  |||| ||
Db  1008 RTTPKVTIVIVSTQNPTTTSKTSTVTI---TTP-NPSPSTQRPTTTTRQTSIFAST 1061

Qy   619 PAPTTPKEPAPTTPTTPKRPAPTTPTPEPPTTSEVSTPPTT-TKE 667
          || ||||  |||  | || |||  |  |  || |  ||  |  ||
Db  1062 TSIGTTRIPTTNPQNSTTSSTDLTTVIRPPCPDPDSTSDKNTNTNACTQEIQQVNIIELQS 1121

Qy   667 PQKQEQFTHTRTHPALTGSRNTLGGQEVPDVMDDAPSSAAEAESGQATTAKAPTMSTLAAA 718
          |                |   |  |             | ||  ||    |  |
Db  1122 --PTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKD 1181

Qy   719 KTTER----DLRTTPETTTTAAPKMTKETATTTEKTTTESKTTATTTQVTSTTTQQTTTPKIT 775
          ||             |   |  | || | | ||  |   |  ||  | | |
Db  1182 HLLQKLFHIISTTPPSREHAP--TQRPSSQPSSSQRSR-GVTIAQMARHNLATSKPFTAH 1238

Qy   776 TLKTT----TLAPKVTTIKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKA-PK 830
          :|:        |  |  |    |  |    | |  |  |  ||  ||  |  |   |
Db  1239 SIRLSIQQLASTQKRSIPPKTLVYHNTTKEPED-SEYDSETSEQYTDEDNEVLDKTQPR 1297
```

```
Qy   831 KPTSTKKPKTMPRVRKPKTTPTTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVE 890
          | ||||      |  |  |  |  |  | |  |:|:  |   | ::  |||  |||| |
Db  1298 AMSSTTVAAVLPAVPSTTTEREPQK-TSSSP--SPT---KATSSTTTQPTIETTTGDLEY 1350

Qy   891 VNPKSED 897
          | |
Db  1351 DSSGSSD 1357
```

```
RESULT   7
ID   O96449            PRELIMINARY;    PRT;    1489 AA.
AC   O96449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RACE 1-11;
RA   Goenhardt B.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061185; AAC72308.1;
SQ   SEQUENCE  1489 AA;  764037 MW;  764CC79D1C2F5163 CRC64;

Query Match          17.0%;  Score 853;  DB 10;  Length 1489;
Best Local Similarity  32.3%;  Pred. No. 1.8e-46;
Matches 355;  Conservative 46;  Mismatches 475;  Indels 222;  Gaps 49;

Qy     9 TKKKPPTPKPPVVDEAGSGLDNGDFKVTTPDTSTT------QHNKVSTSPKITTAKPINPR 62
          ||| | |||| |||||  |:||:
Db   338 TPYAPPEKPVDVEETTYVTEESTTYAPTKSETNAPTERMHYAHIEKPCDTEVTMYAPTEET 397

Qy    63 PSLP---------------------PNSDT---SKETSLTVNKETTVETKETTT--TNKQTSTD 100
          ||                   ::|                  :||:
Db   398 TYAPTEETTYAPTEETTYAPTEETTYAPTEEPTYEPTEETTYAPTEETTYAPTEETTYAP 457

Qy   101 GKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPKPTKAETTTKGPALTTPKEPPTTPTTPKEPA 160
          | |||  |  |:                | | |         |  ||  ||| | |||||
Db   458 TEETTYAPTEETPYAPTEETTYAPTEETTYAPT---EETTYASTEETTYAPTEETTYAPA 514

Qy   161 STTPKEPTPTTIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT--TTKEPAPTTPKEPAPTT 217
          ||| | |   :|||                |  ||    |:|     | |||  || |||
Db   515 EETPYEPTEET-TYAPTEETTYAPTEETTYAPTEETTYAPAEETTYEPTEET 573

Qy   218 TKEPAPTTTKSAPT--------------TPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKE 267
          |||||| | |||              |||  ||  |||  |||  ||||||
Db   574 TYAPTEETT-YAPTEETMYAPTEETTYAPAEETPYEPTEETTYAPTEETTYA 632

Qy   268 P--------------AP--TTKEPAPTTPKEP-----APTAPKKPAPTTPKEPAPTTP 304
          |              ||  |||||||  ||      ||| | ||||| |||| |||
Db   633 PTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEE 692

Qy   305 KEPAPT-----TTKEPSPTTPKEP-----APT--TTKSAPT---------A 339
          ||||||     |||  ||| |||      ||| ||| |||
Db   693 TTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYGPTEETTYA 752

Qy   340 PT-TTKSAPT--TPKEPSPTTPKEP-----APTTKEPAPTTPKKPAPTTPKEPAPTTPKEPAPT-- 388
          || |||  |   |||  |||||         |||| ||| |||  ||| | |||  |||  |||
Db   753 PTEATTYAPTEETPYAPTEETTYEPTEETTYAPTEETTYAPTEETTYAPTEE 812

Qy   389 TKEPAPTTTKPAPTTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPKKPAPTTPEE 448
          ||||||| | ||| |  | |||  |  | |||  |  |||  | |||  |||| |||
Db   813 TPYEPTEETTYAPTEETPYEPTEET--TYTPTEETTYAPTEEKTTYAPTEETT 870

Qy   449 LAPT--TPEEPTPTTPEEP----------APT----------TPKAAAPNTPKEPAPTTPKEP- 489
          |||   |||  | || |||          |||          |||   |||  |||  |||
Db  1239
```

Query Match   16.1%;   Score 806.5;   DB 10;   Length 1315;

```
RESULT    8

ID      Q9SPM0          PRELIMINARY;    PRT;    1315 AA.
AC      Q9SPM0;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE      EXTENSIN-LIKE PROTEIN.
GN      PEX2.
OS      Zea mays (Maize).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC      Panicoideae; Andropogoneae; Zea.
OX      NCBI_TaxID=4577;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=POLLEN;
RA      Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA      Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT      "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL      Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AF159297; AAD59980.1; -.
DR      InterPro; IPR001611; LRR.
DR      InterPro; IPR001998; Xylose_isom.
DR      InterPro; IPR002965; P_rich_extensn.
DR      InterPro; IPR003592; LRR_out.
DR      Pfam; PF05560; LRR; 3.
DR      PRINTS; PR01217; PRICHEXTENSN.
DR      PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR      SMART; SM00370; LRR; 3.
SQ      SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;
```

```
Db    871  YAPTEETPYEPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPT 930
                 | |::||  ||
Qy    490  -----APTTPKEPAPTTPKETAPTTPKGTAPT--TLKEPAPTTPKKPARK-ELAPT--TT 539

Db    931  EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 990
              ||  |  :|  ||  ||  |::||  ||
Qy    540  KEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP------APTTPKGTAPT----- 585

Db    991  YAPTEETMYAPIEETTYAPTEETTYAPEETPYEPTEETTYAPTEETTYAST 1050
              |:  | |  ||  |  |::||  ||
Qy    586  -------TLKEPAPTTPKKP------APKE---LAPT--TTKGPTSTTSDKPAPTTP 624

Db   1051  EETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETT 1110
              |  ::||  ||  ||  |  |::||  ||
Qy    625  KETAPTTPKEPAPTTPKKPAPTTPEPTPPPTTSEVSTPTTTKEPT--TIHKSPDESTPELS 682

Db   1111  YAPAEETPYEPTEETTYAPTEETTYAPTEETTYGPTEETTY-APTEATTYAP 1169
               :|  |:  ||  |  ::||  ||
Qy    683  AEPTPKA-LENSPKEPG-----VPTTKTPAATKPEMTTTAKDKTTERDLRTTP----ETT 732

Db   1170  TEETTYAPTEETTYEPTGTTTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 1229
              || ::||  ||  ::||  ||
Qy    733  TAAPKMTKETATTEKTTESKITATTTQVTSTTTQDTT---------PKITT---LKT 779

Db   1230  TYAP------TEETTYEPTEETTYAPTEETTYAPIDETTYGPTEE 1284
               |  ::
Qy    780  TTLAPKVTTTTKKTITTTEIMNKPEETAPKKPDRAT--NSKAPTTKP-QKPIKAPKKPT--- 833

Db   1285  TTYAPTEATTYAPTEETPYAPTEETTYGETTTYAPTEETTYAPTEETTYAPMEETPYE 1344
              |  |  ||
Qy    834  -------STKKP-KIMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP--N 881

Db   1345  PAEESTSTVSTEKPCNTEEETDEPTDEPTDE--PSDEPTDEPTDLPTDEPSTPCDN 1402
              |:|:  |  |
Qy    882  ----------------------------STKKP----------------------- 899

Db   1403  QGINGIGVENKVRYNNAG 1420
              | :||
Qy   1403  OTPNSKLVEVNPKSEDAG 899
```

Best Local Similarity   25.0%;   Pred. No. 1.5e-43;
Matches 255; Conservative   86; Mismatches 459; Indels 219; Gaps 33;

```
Qy     13  PTP--------KPPVVDEACSGLDNGDFKVTTPDCSTTQHNKVSTSPKITTAKPINPR  62
              |||   |   |  ::|
Db    923  LAPISSPPSSEPKSPSSPPMVEKTSPPPATVSSPPPTLKSSSPPPP--APVSSPPPVVKSSP 979

Qy     63  PSLPP-------------NSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQ 112
              |||
Db    412  PIPGGGPPSSSVPGKPPSVPG--------KPAAAPAMPTTPHTTPPDVSP----EPL-PE 455

Qy    159  PASTTPKEPTPTTIKSAPTTPKEPAPTTPKE-KEPAP------TTTKEPAPTTP 210
              ||||   |  :  | |::||  ||
Db    572  PPMKSP--PPPARVASPPPLMKSPPPPAPVASPPPQPLKSPPPVVLWLSTPSVKSPPPVP 629

Qy    211  KEPAPTTPKEPAPTTPKSAPTTPKE-----PAP--TTP-----KKPAPTTPKEPAPTTPKE- 259
              || ::||  ||
Db    512  SPPATPVKSSPPPAAVVLPPPAKTSPPPEAPVASPPPAPVSSPPQPVKSPPPPAPVASPP 571

Qy    260  PTPTTPKEPAPTTPKEPAPTTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTT 319
              |||   ||  |::||  ||
Db    159  PASTTPKEPTPTTIKSAPTTPKEPAPTTPKE-KEPAP---------TTTKEPAPTTP 210
```

RESULT   9

Db   1272 -------------PKKEEDSTA-------

```
Q20007
ID   Q20007;     PRELIMINARY;     PRT;   1274 AA.
AC   Q20007;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   COSMID F35A5.
DE   F35A5.1.
GN   F35A5.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Leimbach D.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U46675; AAB52641.1; -.
SQ   SEQUENCE 1274 AA; 138065 MW; 07F6B0D292C1799F CRC64;
     "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
     elegans.";

Query Match      16.0%;  Score 802;  DB 5;  Length 1274;
Best Local Similarity   28.0%;  Pred. No. 2.7e-43;
Matches  290;  Conservative 102;  Mismatches 414;  Indels 230;  Gaps 56;

Qy      6 KNRTKK-KP-----TP------KPPVDEAGSGLDNGD-----FKVTPDTSTTQHNKVSTS 50
            ||| ||  || |       :  :     |||
Db    274 KNPTKKKPKPPWEDEDTPVEEVKEPPVPEKKAPVLKKKDPAPAAKARDPSPSKAAPKVEPS 333

Qy     51 PKITTAKPI-NPRSLPPNSDTSKEISLTVNKETTVETKETTTNKQTSTDGKEKTTSAK 109
          | |||  | |||:|  |  |::||    |   | |  |:   |   |||
Db    334 SPVVPPTPVKNPVKKYKPPWEVDDEPAEEVKKPSAPEKKTPVLKRKEPEPSSTTPSSDPS 393

Qy    110 ETQSIEKTSAKDLLAPTSKVLAKPTPKAE------TTTTKGPA 148
          :   ||  |    || ||||||:  |::           LTTP
Db    394 PKKAAPAVKPRDSSPKKATPLQADPKAQEVPPTPVKNPVKKYKPPWEVDDEDPVEEVKQP 453

Qy    149 KEPPTPTTP-----KEPASTTPKEP----TP-TTTKSAPTTPKEPAPTTTTKSAPTTPKEP 199
             |  |       :|         |:        |     |||| ::|
Db    454 EAPAKKTPVLKRKEPAAKDTAKPAIKSKPETPEKKDPVKPRDSSPKKVAAKPDSAQAPA- 512

Qy    200 TTTTKEPA------------PTTPKEPAPTTTKEPAPT-------TTKSAPTTP 233
                   |||  ::||         |:||
Db    513 TPVKNPVKKWKRPPWEDDETPADDVSKPIDAKKTPSLAKKDPAPAKESLKPKADITKAPAKP 572

Qy    234 KEP-------APTTPKPAPTTPKE----PAPTTPKEPTPTTPKEPAPTTKEPAPTTPK-EP 283
                     ||: |  |  :   |   |       |  |:| |
Db    573 RDPSPKKVAPTAPEKKIPVLAKKEPAGPADSKTKEPEKSKPRDPSPKKAVPAKPVPKTEV 632

Qy    284 APTAPKKPAP------TTPKEPAPTTPKEPAPTTPKEP-----SPTTPKEPAPTTPKSAPT 333
          | ||| |         |||| |||| |||  ||   |         |  |||| ||   |
Db    633 APAAVKPEPISKPKDTAPKKAEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPE 691

Qy    334 TTPKEPAPTTPKSAPTTP-KEPSPTTTPKEPAPTTPK--EPAPTTPKKPAP--TTPKEPAPT 388
             |  |    |  |    |:|     |    |   |  :|   | |   |    |  |
Db    692 PEKK-TPVLAKKAPTKPDSEAAADPVSGPSSKDPKLAKKAPVKPRDPSPMKAVPIKPAPK 750

Qy    389 TPKEPAPTTTKKPAPTA-PKEPAPTTPKETAPTTPKKLITPTTP-----EKLAPTTPEKPA 442
           |  |        |   ||| |  |    |  |          |       ||  | |
Db    751 T--EVPPAVVKKPEPVAKSRDPSPKKAK-AEPNSP--VVPPTPVKNPVKKWKPPWEDDDA 805

Qy    443 PTTPEELAPTTPEEP------TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTT-PKE 496
           |||   |   |          |  |   | |  |       |:    |   |
Db    806 PAEPVNVPEPEKKTPVLAKKTPVKPRDPSPKKAVPAKPSTKIDAPPVSVKKPEPVSKPKE 865

Qy    497 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPT-TTKEPTSTTSDKPAPTTP 555
          | |       |   |       |  | |:|     | |
Db    866 PSPKKAEPNSPVVP----PTPVKNPVKKW-KPPWEDDDEPTEEVKKIPSE--PEKKTPVLA 918

Qy    556 KGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPT------TTKGPA------KEL 603
             | ||     |    ||  |   |          |          |         |
Db    919 K-KEPEKPRD-APKVAAKPRDPSPKKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEA 974

Qy    604 APTTTKGPTS-------TTSDKP-----APTTPKET-------APTTPKEPAPTTPKKPAPT 646
          | |      |        |  |       | |  |:        | | |     |
Db    975 PPTPVKNPVKKWKPPWEDDDEPAEPVSAPEPEKKTPVLAKKAPAKPRDP---SPKKAAPV 1031

Qy    647 TPETPPPTTSEVSTPTTTKEPTTIIHKSP----DESTPELSA-EP---TPKALENSPKEPG 698
           |   |        |  | |          |  |   |   |       |  ||  |
Db   1032 AAK-PDPKIPEV-PPTPVKNPVKKWKPPWEDDDEPSEPVSAPEPEKKIPVLAKKAPTKPA 1089

Qy    699 V-------------PTTKIPAAT-----KPEMTTTAKDKTTTERDLRTTPETT--TAAPK 737
                       |  | ||     | |:     |   | |||| |:| || |
Db   1090 TKPDSEAAADPVSGPTSKDPKLSKKAPVEKPKPTTDPKDDKLKPSPAKKPEKAPEPAAPK 1149

Qy    738 MTKEFATTTEKTTESKIIATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTTITTE 797
          : |    |:||  |   |  ||:|  | | | | ||  | | || | |  |   | | |
Db   1150 KWKPVWDDDDEPEADFTVPAPSKKPDTEDPADPLG--------GPKTKDPK----- 1193

Qy    798 IMNKPEETAKPKDRATNSKKATTPKPQKPTKAPKKPTSTKKPKIMPRVRKPK-------- 848
          :  |||   | :        | |  | |       |:  |  ||   |
Db   1194 -LNKRAPAEKPTEK------PKPKEVSKEPPKPKPTEPPKP-AAPKKWKPPWEDDDEPE 1243

Qy    849 ---TPTPPRKMTSTMP 861
             |  |  |: |  |
Db   1244 ADFTMPAPKKPDTEDP 1259
```

```
RESULT 10
Q9LIE8
ID   Q9LIE8     PRELIMINARY;     PRT;   1480 AA.
AC   Q9LIE8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
```

RX   PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP001306; BAB03062.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   1480 AA; 147153 MW; D1ACOC79F155E732 CRC64;

Query Match   16.0%; Score 800.5; DB 10; Length 1480;
Best Local Similarity   29.6%; Pred. No. 3.9e-43;
Matches 307; Conservative 54; Mismatches 505; Indels 171; Gaps 45;

```
Qy     10 KKKPPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTS-----PKITTAKPINPRPSL 65
          | | |||| || | ||||                            ||:  ||||  |:
Db    117 KPRPHPKPPNVK-----------PHPHPKPPTKPHPHPKPPTKHHPHPKPPTIKPPPKPPSV 167

Qy     66 PPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTT 125
          |                                               |: | ||| | |
Db    168 KPPPSTPK----------------PPTTNPPPSTPQPPPHKPPPCTPTPPPVASPPMATPP 211

Qy    126 SKV--LAKPTPKAETTIKGPALTTPKEPPT---TTP--KEPASTTPKEPTPTTIKSAPTTP 179
          :::   :|:| |  |  || |  | :|        |:  | |:|||  ||||| ||||  ||
Db    212 TQMPPIATP-PIAKSPVATPPIANPPTATPPITPPIINPPVATPPITTPPIANPPIATP 270

Qy    168 KPPPSTPK--------------PPTTNPPPSTPQPPPHKPPPCTPTPPPVASPPMATPP 211

Qy    180 KEPAPTTT----KSAPTTPKEPAPTTT-----------------KEPAPTTPKEPAPTTTKE 220
          ||||                |||| |||                 | ||||||||||||| |
Db    271 PPTTTPPIAKPPIATPPISTPPAATPPAATPPITTLP 326

Qy    212 TQMPPIATP-PIAKSPVATPPIANPPTATPPITPPIINPPVATPPITTPPIANPPIATP 270

Qy    221 PA----------PTTTKSAPTTP-KEPAPTTPKKPAP-----TTPKEPAPTTPKEPTPTTPKEPAP 270
          |:           |||| || | |||||  ||                | |:|||||   |||||
Db    327 PAKPPVAISPIVT---PPVTPIAQPPVATPTATPPVATPPIATPPTSKSPISTPPISESP 384

Qy    271 TTIKEPAPTTP-KEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTT----KEPSPTTPKEPAP 325

Qy    326 TTTTKSAPTTTKEPAPTTTKSAPTTTKEP---SPTTTKEPAPTTTPKEPAP---TTPKKPAPT 380

Qy    271 PVAAPPITNPPISKPPVUTP------PPTTTPPIAKPPIATPPISTPPAATPPAATPPITTLP 326

Qy    445 TAI----PPVAKPVETPIPTAIPPPAKPPISTPPISKPPVAIPPAATPPITTPPVKPPVA 501

Qy    385 VATPPTATPSPIKTPPPAKPPVATPPIAKSPIATPPTATPPVATPPIEKPPVATPPTTTP 444

Qy    381 TPKEPAPTTPKEP--APTTTKKPAPTAP--KEPAPTTPKETAPTT------PKKLTP--T 428

Qy    502 TPPLAIPPVAKPPVVTPPATTPPIATPPIAKSPVATPPTATPPVATPPIAKPPVVTPPTT 561

Qy    429 TPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPRKAAAPNT---PKEPAPT 484

Qy    562 TPPTATPPVAKPPVATPIATPPPKAKPPISTPPISKSPVATPPAATPPITTPPPAKPPVA 621

Qy    485 TPKEPAPTTPKEPAPTTPKETAP---TTPKGTAPTT---LKEPAPTTPKPAPKELAPTTT 539

Qy    622 TPPIATPPIAKPPVATPPTATPPIATSPVAKPPVAIPPTPPPAKPPVAIPPIATPPPVA 681

Qy    540 KEPTSTTSDKPAPTTPKGTAP--TTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKP 598

Qy    682 KPPVAT----PPTATPPIATPPIATPPVVTPPIATPPVATPPIAKPPTTT--PPTATPPVA 736

Qy    599 APKELAPTTTKGPPST----TSDKPAPTTPKETAPT--TPKEPAPTTPKKPAPTTP-----E 649

Qy    737 MPPIATPPPAKPPIATPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVA 796

Qy    650 TPPPTTSEVSTPTTTKEPTTIH--KSPDESTPELSAEPPKALENSPKEPGVPTTKTPAA 707

Qy    797 TPPIAKPVATPPTATPPIAKPPILTPPISKPPVATPPAA 856

Qy    708 TKPEMT--TTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTT 766

Qy    857 TPPITTPPPAKPPVATPPIATPP--IAKPVATPPTATPPIATSPVAKPPVAIPPIKTP 914

Qy    767 QDTTPFKITTLKTTLAPKVTTTKKT----IITTEIMNKPEETAKPKDRATNSKAITPKP 822
```

RESULT 11
Q41805

ID   Q41805;   PRELIMINARY;   PRT;   1188 AA.
AC   Q41805;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z34465; CAA84230.1; -.
DR   Mendel; 14346; Zeama; 2368; 14346.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   SMART; SM00370; LRR; 4.
KW   Signal.
FT   SIGNAL.
FT   SIGNAL   1   27   POTENTIAL.
SQ   SEQUENCE   1188 AA; 120980 MW; 2C77C7FBD7130149 CRC64;

Query Match   15.6%; Score 784; DB 10; Length 1188;
Best Local Similarity   29.9%; Pred. No. 3.6e-42;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

```
Qy     13 PTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKIKTTAKPINPRPSL----PPN 68
          |||                                       ||  ||||||     |||
Db    460 PTPHSPPAD-------DYVPPTPVPGKSPPATSPSPQVQPPAASTPPPSIVKLSPPQ 510

Qy     69 SDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV 128

Qy    511 APVG-----------------------------SPPPPVKTTSPPAPIG----SPSPPPVSVV 541

Qy    129 -----LAKPTPKAETTTKGPALTTTPKEPTTPKEPTPTTTKEPPTTTIKSAPTTP-KEPAPTT 183

Qy    542 SPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVASPPPPVKSP--PPPTLVASPPPPVKSPP 599

Qy    184 PTTTKSAPTTP---KEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTT-KEPAPTT 240

Qy    600 PPAPVASPPPPVKSPPPPTPVASPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPP 659

Qy    241 PKKPAPTTPKEPAPTT------PKE---PTPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKP 291

Qy    660 PAKSTPPPEEYTPPPSVKSSPPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP-PSSPEKP 718

Qy    292 APTTPKEPAPTTPKEPAPTTTKEPSPTTTPKEPAPTTTPKEPAPTTTKSAPTTTKEPAPTTK 351

Qy    719 SP--PKEPVSSPPQTP----KSSPPPAPVSSPPPPIPVSSPPALAPVSSPPSVKSSP---- 768
```

```
Qy    352  EPSPTTTKEPAPTTKEPAPTTTKKPAPTTKEPAPTTTKKPAPTTKEPAPTTTKKPAPTTKEPAPTTPKPAPKEPAP 411
Db    769  PPAPLSSPPAPQVKSSPPPAVSSPP-PPAPKSSPPLAPVSSPPQVEKTSPPPAPLSSPP 827

Qy    412  TTPKETAP---------TTP----KKLTPTTPEKLAPTTPEKPAP-------TTPEELAPTTPEEPT 458
Db    828  LAPKSSPPHVVSSPPPVVKSSPPPAVSSPPLTPKPASSPPAHVSSPPEVVKPSTP--PA 885

Qy    459  PTTPEEPAPTTPKAAAPNTPKE-PAPTTPKEPAPTTPKETAPTTPKGTAPTT 517
Db    886  PTTVISP-PSEPKSSPPPPPVSLPPPIVKSSPPPAMVSSP-PMTPKSSPPPVVVSSPPPT 943

Qy    518  LKE--------PA-----PTTPKK----PAPKELAPTTTKE---PTSTTSDKPAPTTPKGTAPTT 563
Db    944  VKSSPPPAPVSSPPATPKSSPPPAPVNLPPPEVKSSPPPPIVSSPPPA---PKSSPPPAP 1000

Qy    564  -KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTPKKPAPKELAPTTTKGPTSTTSDKPAPT 622
Db   1001  MSSPPPPEVKSPPPAPVSSPPPPAPVS---SPPPPVKSPPPPAPAVSSP----PPVKSPPPPAPVSSPPPPV 1057

Qy    623  -TPKETAPTTPKEPAPTTPKKPAPTTPEIPPPTTSEVSTPTTTKEPTTTHKSPDESTPEL 681
Db   1058  KSPPPPAPISSPPPVVKSPPPPAPAPVS--SPPPPVKSPPPPAPAVSSPPPPIKSPPPPAPAVS 1115

Qy    682  SAEPTTPKALENSRKEPGVPTTKTPAAATKPEMTTTAKDKTTTERDLRTTTPET 731
Db   1116  SPPPAP------VKPPSLP-PPAPVSSPPPVVTPAPPKKEEQSLPPPAES 1158

RESULT  12
ID    P70670          PRELIMINARY;     PRT;   2187 AA.
AC    P70670;
DT    01-FEB-1997 (TrEMBLrel. 02, Created)
DT    01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE    MUSCLE-SPECIFIC FORM GP220).
GN    NACA.
OS    Mus musculus (Mouse).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX    NCBI_TaxID=10090;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=9312450; PubMed=8698236;
RA    Yotov W.V., St-Arnaud R.;
RT    "Differential splicing-in of a proline-rich exon converts alphaNAC
RT    into a muscle-specific transcription factor.";
RL    Genes Dev. 10:1763-1772(1996).
DR    EMBL; U48364; AAB18734.1; -.
DR    EMBL; U48363; AAB18732.1; -.
DR    MGD; MGI:106095; Naca.
DR    InterPro; IPR002715; NAC.
DR    InterPro; IPR003037; TS-N.
DR    Pfam; PF01849; NAC; 1.
DR    Pfam; PF02094; TS-N; 1.
SQ    SEQUENCE  2187 AA;  003464AA864DEBFD CRC64;

Query Match             15.5%;  Score 776;  DB 11;  Length 2187;
Best Local Similarity   27.7%;  Pred. No. 1.9e-41;
Matches   288;  Conservative   122;  Mismatches   389;  Indels   240;  Gaps   49;

Qy     34  VTTPDTSTTQHNKVSTSPKTTTAKPINPRRSLPPNSDTSKE----------TSLTVNKETT 84
Db    806  VQSPKVDPIMSDVPTSPKKTSATAV-PK--------DTSATLSKSVPAVTSLSPPKAPV 857

Qy     85  VETKETTTTINKQTSTDGEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPA 144
Db    858  APSNEATIVPTEIPSLKNALAAATPKETLIATSIPKVTSPS----PQKTPKSVSLKGAPA 913

Qy    145  LTT--------------------------PKE-------PTPTTPKEPASTT-------------PKEPTPTT 171
```

```
Db    914  MTSKKATEIAASKDVSPSQPPKEVPLLQHVPPTSPPKSPVSDTLSGALTSPPPKGP-PAT 972

Qy    172  IKSAPTTPKE-PAPTTTKSAPTTPK------------------EPAPTTTKEPAPTTTKEP 221
Db    973  LAETPTYPKKSSPKPAASKKTPATPSPEGVTAVPLEIPPCSKKKAPKTAAPKESSATSSSKR 1032

Qy    222  APTT---TKSAPT--------------------TPKEPAPTTPKK--PAPTTPKEPAPTTP 257
Db   1033  APKTAVSKEIPSKGVTAVPLEISLPLKETSKSATPGEKSASSPKRSPKTGAGPKE----TPP 1089

Qy    258  KEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKKPAPTT---PKEPAP---TTPKEPAPTT 311
Db   1090  GGVTAVPPEISLPPKETPQNATPNESLAASSQKRSPKTSVPKETPPGGVTAMPLEIPSAP 1149

Qy    312  TKEPSPTTPK-------EPAPTTTKSAPTTTPKEPAPTTPK-------KSAPTTPK-----EPSPTT 357
Db   1150  QKAPKTAVPKQIPTPEDAVIILAGSPLSPKKASKTAAPKEAPATPSVGVIAVSGEISPSP 1209

Qy    358  TKEPAPTTPKEPAPTTPKETAPTTPKKKPAPTT---PKE-------PAPTTPKEPAP------T 396
Db   1210  KKTSKTAAPKENSATLPPKRSPKTAAPKETPATSSEGVTAVPSEISPSPTPASKGVPVT 1269

Qy    397  VGPKETSLEGAPAVPLEIPPSHKKAAPKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-A 455
Db   1270  LTPKGAPNALAE-SPASPKVPKTAAPEEITSTTP-----SPQKIPKVAGPKEASATPPS 1322

Qy    456  EPTPTT---------PEEPAPTTPKAAAPNTPKEPAPTTPKE-PAPTTPKEPAPTTPKETA 506
Db   1323  KKTPKTAVPKETSAPSEGVTAVPLEIPPSPRKAPKTAAPKETPAPS-PEGATTAPVQIP 1380

Qy    507  PTTPKGTAPTTLKEPAPTTP----------------KKPAPKELAPTTTKEPTS- 544
Db   1381  PSPRKGSKKAGSKE-TPTTPSPEGVTAAPLEIPISSKKTSRMASPKETLVTPSSKKLSQT 1439

Qy    545  ------TTISDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKKTAAPKTAAPKESSATISSKKT 599
Db   1440  T--SLAQTAPPSLQKAPSTTIPKENLAAPAV----LPVSSKSPAAPARASASLSPATAAP 1496

Qy    600  PKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKEPAPTTPKKPAPTTPETP- 654
Db   1497  PKTAAPPSER--VTTVPPEKPA-TPQKASGTTIASKVPVPAETQEVAVSSRETPVTAVPPV 1554

Qy    655  TSEVSTPTTTK-----EPTTIHKSPDESTPELSAEPTTPKALENSPKE-PGVPTTKTPAA 707
Db   1555  KNPSSHHKTKSKTIELKEAPATLPPSPTKSPRKIPSSKKAPRT--SAPKEFPASPSIK-PVT 1611

Qy    708  TKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVISTTTQ 767
Db   1612  T--SLAQTAPPSLQKAPSTTIPKENLAAPAV----LPVSSKSPAAPARASASLSPATAAP 1665

Qy    768  DTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTK 827
Db   1666  QTAPKEATTIPSCKKAAATETPIEISTAPSLEGAPKETSE----TSVSKVLMSSP----- 1716

Qy    828  APKKPPTSTKKKPRKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSK 887
Db   1717  -PKKASSSKRASTLP------------------A-TTLPSLKEASVLS------PIATSSGK 1752

Qy    888  LVEVNPKSEDAGGAEGETP 906
Db   1753  DSHISPVS-DACSTGTTTP 1770

RESULT  13
ID    Q9VEL9          PRELIMINARY;     PRT;   2112 AA.
AC    Q9VEL9;
DT    01-MAY-2000 (TrEMBLrel. 13, Created)
DT    01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT    01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE    CG4090 PROTEIN.
GN    CG4090.
```

```
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nixon K.A., Nusskern D.R., Pacleb J.M., Palazzolo M.,
RA   Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003716; AAF55402.1;
DR   FlyBase; FBgn0038492; CG4090.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF01607; Chitin_bind_2; 10.
DR   SMART; SM00494; ChtBD2; 11.
DR   PROSITE; PS00022; EGF_1; 1.
SQ   SEQUENCE 2112 AA; B91018E551A5D36 CRC64;

Query Match             14.7%;  Score 739;  DB 5;  Length 2112;
Best Local Similarity   27.2%;  Pred. No. 4.2e-39;
Matches  281;  Conservative  98;  Mismatches  339;  Indels  316;  Gaps  44;

Qy      4 NKKRTKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTQHNKVSTSPKITTAKPINPRP 63
            :|| :||  || |  :|  || :||: |||   | |::||  |||| |  ||:|
Db   1023 NHPDQIQTRPLKCKKVVSQGGSSSN--------STSNSSSSSNNSGSSSNSGSSSS---- 1070

Qy     64 SLPPNSDTSKETSLTVNKETTV---ETKETTTTNKQTS----TDGKEKTTSAKETQSI- 114
           :||: ||| || : |  |:|       |:|| |||| ||||   ||  ::|||:| |:
Db   1071 SSSSNSGSSSNTGSSSNSGCASSSGGSSSNQGSSSNSGGSSSNSGSGNQSTTSTSSSSSS 1130

Qy    115 ----EKTSAKDLAPTSKVLAKPTPKAETTTKGP--- 143
               |:|: |||:||  | || ::|  ||:||||
Db   1131 NNNNQGSSSSSSSSSSSSSTSSSKPNPSETCKVNGQFIGDRSDCAKFYRCVDNDRGGFNMVPF 1190

Qy    144 -----------------------------------ALTTPKEPTPTTPKEPASTTPKEPTTPTTIKSAPTT 178
                                               :|| |||| :| ::|||  | | ::||
Db   1191 SCGPGTVWDAQMQACNHAWAVKECGGIAPPTTSPPTTSR-PTTASTSRPSDQTSTSRPTG 1249

Qy    179 PKEPAPTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTT 217
          | :|   |   ::| |
Db   1250 P-----PTTARPVTARPTTSSPTTASSSQTTSPVTQAPNTDGKCRSEGFMADPNNCSKFYR 1305

Qy    218 ---------TKEP----APT-- 248
                    |||     ||
Db   1306 CVRNNKGGFTSIPFQCGAGTVWDQDLQTCNHNFNNCSTGTESTTPKPPC--EPAINGTTA 1363

Qy    249 PKEPAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTTPAPKKPAPTTPKEPAPTTPKEPA 308
          | :|  ||| | :|||  :||  |: || |  :|||  ||  |  ||| ::|||  |  |||
Db   1364 TSTSSTTTP--PPTTTDLPPTSTTGLP-PTTTTELPPT--------------TTTDLPPTTTTRLPP 1413

Qy    309 PTTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKE---EPSPTTTKEPAP-- 363
          ||||  | |||    |:|||:  |:||| | |||    |  ::|     :|| ||| |||
Db   1414 TTTTSLPPTTTTGLPPTTTTGAQPTTTTTLSSETETSTVTTSPESTTQPPSTTTMKPLPAG 1473

Qy    364 ---------------TTPK------------- 373
                         ||
Db   1474 TECTGEGYMADPEDCRKYFRCINAGASYRKYNFTCPKGTGWNEEKVQTCDYVENIPRCSKL 1533

Qy    374 ---KEPAPTTPKE-PAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPT 428
             |:|   |||   ||:|| ||| | |:|  |  ||:| |||||:  ||
Db   1534 PAEPITTTTPSEESKDPGSTTPQSTDEPTTVTKPI-TKPTE-EPSTEKPQKPTTQYPEKPT 1591

Qy    429 TPEKLAPTTPEKPAPTTPEELAPTTPEEPPTTTPEEPAPTTPKAAAPNTPKEPAPTTPKE 488
          |||  | |||:||  ||| :  ||||||  ||||||  ||| |:  ||||||  |||||
Db   1592 TTEE--PEKPQKPTT-TTEYPQKPTTTEEPTPEKPQKPT-TTEYPQKPTTTEEPTPEKPQK 1647

Qy    489 PAPTT-PKEPAPTTPKETA----------PTTPK--GTAPTTLKEPAPTTPKKPAPKELAPTTTK 540
          | ||| ||:|  || |       |       || |  | ||| ||  |||  ||| :| ||||
Db   1648 PTTTEYPQKPTTEEPTTSIPGYNPTTTSVPGYNPTTTPIPVETTTTSTPGYK--PTTTG 1704

Qy    541 EPTSTTTSDKPAPTTPKGTAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKK-- 597
          ||  || |  | |||||  | |||   |  |||   | |||  |  ||
Db   1705 EPITTTT-LPSTTTDAIQEPTTSKKPEPTTTESPESSTPEGSVTTLQPEPQPNNCSSE 1763

Qy    598 ---PAPKELA------PTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTP 648
             | |:         | ||   ||  | |  | |||:| ||||||  ||| |||  ||
Db   1764 GFFPDPEDCSRYYRCVDAAKNGKYQVYAFKCGKGTVWDTSTETCNYADQVSGN------ 1816

Qy    649 ETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAT 708
            | ||| ||| ||| :| ||   ||:||| |   |||   |:| ||| || |||
Db   1817 ----CSSGQTTTPGTTTEPGTTESTTTSSSGKP----ETTSKAPEN--------TTTWA-- 1857

Qy    709 KPEMTTTAKDKTTERDLRTTPETTTPAAPKMTKETATTTEKTTESKITATTTQVTSTTTQD 768
           |: |||     :  | :|||| |:  |  :| |  ||| |:|  | |||| ||||||:
Db   1858 -PETTTT-----VASETTTTTSGTT-------TTTATPETTTKP 1895

Qy    769 TTPFKTTTLKTTTLAPKVTTTKKTITTTEINKPEETAKPKDRATNSKATTPKPQKPTKA 828
          ||   |||  |||     ||||| |||||
Db   1896 PKP--------ETTTIAGEETSTSKSPTTTE-----------SPAPSTNTSA 1928

Qy    829 PKKPISTKKPKTMP 842
          | |  |   |  :|
Db   1929 P-------CPETGP 1935
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang Y., Moreno C., Singh M.;
RT   "Cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
     EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
     InterPro; IPR003882; Pistil_extensn.
     InterPro; IPR0029965; P_rich_extensn.
DR   PRINTS; PR01222; ATROPHIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSN.
SQ   SEQUENCE   763 AA;   75034 MW;   39168EC45A5916F8 CRC64;

Query Match             14.3%;   Score 715;  DB 2;  Length 763;
Best Local Similarity   28.2%;   Pred. No. 5.8e-38;
Matches  248;   Conservative  39;   Mismatches 361;   Indels 232;   Gaps  39;

Qy   132 PTPKAETTTKGPALTTPKEPTTPTPKEPASTTPKEPTPTTPTTIKSAPTTPKEPAPTTTKSAP 191
Db     3 PVP------APRALALPLPPAPPAPAPEPEKSKPPPFPAP---------------PAPPCWMLVSAAP 46

Qy   192 TTPKEPAPTTPKEPAPTTPTPKEPAPTTTKSAPTTPKEPAPTTTPKKPAPTTPKE 251
Db    47 PCP--PAP----PAPPKPKSKAPPPVPPAPPARELAPPLP--PAP------PEAPRE 90

Qy   252 PAPTTPKEPTTPTT---PKEPAPTTPKEPAPTTPK-EPAPTTAPK-KPAPTTPKEPAPTTPK 305
Db    91 SRPALPPCPPPPVVIPDPEPPAAPVPPAPNSPPPPFPAPPKEVPAPPVP--PVPNSPP 148

Qy   306 EP----APTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAP 363
Db   149 FPPEPPAALNPPAP------PAPPLANSPPLPPPAPPTPAGT--PPAAPWPPVPAAPKSKPA 201

Qy   364 TTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAP-------TTTKKPAPTTPKEPAPTTPKETAP 419
Db   202 SPRRPPAP-----PMPATMEFFPLPPVPPVEPDDPISKETPPAPPAAPPIPPAPVPVPPVPLP 256

Qy   420 TTPKKLTTPTTPEKLAP--------TTPKRPAPTTPPEELAPTTPEEPTPTTPEEPAPTT-- 469
Db   257 PVPNKIPPAPP---APPVAVAVAVLVAPCPLPPLPNNHPAPPRAAPVPGVPLAPLPNSHP 313

Qy   470 ---PKAAAPNTPKEPAP 483
Db   314 PAPPSAPVPGVPLAPLPISGRPVSVWKGSFTTLSTFCCRVCSGEVLAGALNPSRPSRSPL 373

Qy   484 --TTPKEPAPTTPKEPAPTTPKETA------PTTPKGTA--PTTLKEPAPTTP--KKPARKE 533
Db   374 TTTTPALPAPIPPLPPLPINPAVPPIPPLPPVTALAPPLPPIALPLPISPGVPPAP-- 431

Qy   534 LAPTTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPA------PTTPKEPAPTTPKGTAPTTTL 587
Db   432 -PIPPGKPWTTPPLAPAPPEPK-TVPVLPPGPGSCPPSEKPNPPAPPEPEPPEPERKSSPALPP 488

Qy   588 KEPAPTTP---KKPAPKELAPTTTKGP-TSTTSDKPAPTTPKEI--APTTPKEPAPTTP 640
Db   489 APPAPSMPSAVRVPSPSPIPPAPPAPAPRASMPALPPAPPSPPPAPATRLCPPLPPSPPAPNSP 548

Qy   641 KKPAPTTPTPTPPPPTPPPTSEVSTPTTTKEPTTIHKSP----DESTPELSAEPTPKA-LENSP 694
Db   549 --PAPPAAPTTPPKILS--ANPPCPCPVPPAPNRPPAPPAPAPELPAPPDPPPPVANSP 604

Qy   695 KEPGVPTTKTPAATKPEMTTTAKDKTTFRDLRTTPETTTPAARKMTKEFATTTEKTTESKI 754
Db   605 --PAPPAPPSALPFVNPPA------PPTPAAPK---------SRPAL 637

Qy   755 TATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKITTTTELMNKPEETAKPKDRAIN 814
Db   638 PAAPPAPPAPPVRATTP---------PPAPPAPPAPN 665
```

```
RESULT 15
Q9FPQ6
ID   Q9FPQ6      PRELIMINARY;      PRT;     555 AA.
AC   Q9FPQ6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VEGETATIVE CELL WALL PROTEIN GP1.
GN   GP1.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA   Goodenough U.W.;
RT   "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT   plant hydroxyproline-rich glycoproteins.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
     EMBL; AF309494; AAG45420.1; -.
     Interpro; IPR003882; Pistil_extensin.
     Interpro; IPR0029965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSN.
SQ   SEQUENCE   555 AA;   6A584A90465502F5 CRC64;

Query Match             14.1%;   Score 707.5;  DB 10;  Length 555;
Best Local Similarity   28.9%;   Pred. No. 1.3e-37;
Matches  189;   Conservative  66;   Mismatches 243;   Indels 157;   Gaps  22;

Qy   815 SKATTPKPQKP----TKAPKKPTSTKKPKIMPRVRKPKTTPTPRKMTTSTM-----PELNP 865
Db   666 SMALPPAPPDDPIPLLATPPAPPAPPLPMSPPAPPLPPAADPPAPPLTINQPPSPPLAP 725

Qy   866 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET 905
Db   726 VPGAPLAPLPINGRPVPARKNSLI------GSSSGDT 756

Qy   194 PKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTTPKEPA 253
Db    40 PSPAPPS---PAPPSPAPPSP------APPSPAPPSPGPPSPAPPSPPS 80

Qy   254 PTTPKEPTPTTPKEPAPTTPKEPAPTTTKKPAPTTAPKKPAPTTPKEPAPTTPKEPAPTTTK 313
Db    81 P-APPSPAPPSPAPPSPAPPSPAPPSP-APPSPAPPSPSPAPPSPAPPSPSPAPPSP--- 134

Qy   314 EPSPTTPKEPAPTTTKSAPTTTKEPAPTTPKEPSPTTTKEPAPTTPKEPAPTT 373
Db   135 -PSPSPP-----------------------------SPAPPLPPSPAPPS 154

Qy   374 PKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTTKEPAPTTPKETAPTTPKKLTPTTPEKL 433
Db   155 P-APPSPAPPSPAPPSPPVPPSPSPAPPSP-------PSPSPPVPSPAPPSP---- 194

Qy   434 APTTPKEPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTT 493
Db   195 APPVPPSPAPPSP---APPVPPSPAPPSPAPPSP------PPSPSPPSPAPPSPAPPSPA 250

Qy   494 PKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTTKEPSPTTSDKPAPT 553
Db   251 PPSPAPPSPRPPAP----PPPPSPPPPPPRPPPFANTMPPSPSPPPPSPA 298

Qy   554 TPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKPAPKELAPTTTKGPTS 613
Db   299 PPTPPTPPPSPSPPSP-VPPSPAPVPPSPAPS----PAPSPPPSPAP----PTPSPSPSP 349

Qy   614 TTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETTPPPTTSEVSTPTTTKEPTIHKS 673
Db   350 SPSPSPSPS------------PSPSPSPSPIPSPSPSPKPSPSPSPVAVKLVWADDAI--- 393
```

```
Qy   674  PDESTPELSAEPT-PKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTPETT  732
              : :|:    | | : |              |:|: : :  :         |
Db   394  ---AFDDLNGTSTRPGSASRMVGEPDIAGTKCKGNLKGWMPKPSRNPRWGQAVFSGGRTV  450

Qy   733  TAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTK--  790
              : |: | ||               |     | :|  || : : |
Db   451  GSVANVTIRVAFATEK--PALIYSSIELVVYNTG---------ATLIRVPIAANVTRSQIR  500

Qy   791  ----KTITTTEIMNKPEETAKPKDRATNSKATT-PKPQKPTKAPKKPTSTKKPKT  840
                | | |                         | | |   ||||
Db   501  CPGFLTYGTTPIAGYP---------TGIDATTWPNWKIAGVRINMGAGNKKPKT  545
```

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: April 26, 2002, 16:28:53 ; Search time 210.7 Seconds
(without alignments)
946.223 Million cell updates/sec

Title: AA2
Perfect score: 7276
Sequence: 1 MAWKTLPIYLLLLLSVFVIQ.........ARAITTRSGQTLSKVVWVNCP 1363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mhc:*
7: sp_mammal:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7245.5	99.6	1404	4	Q92954	homo sapien
2	7216.5	99.2	1404	4	Q9BX49	homo sapien
3	3716.5	51.1	1054	11	Q9JM99	mus musculu
4	1713.5	23.6	401	6	O77765	bos taurus
5	982.5	13.5	1079	5	Q9N4S7	caenorhabdi
6	946.5	13.0	1049	5	Q9I7S1	drosophila
7	892	12.3	1795	5	Q9I7S1	drosophila
8	863.5	11.9	1489	10	O76894	drosophila
9	846	11.6	1274	10	O96449	phytophthor
10	817	11.2	1480	10	Q20007	caenorhabdi
11	816	11.2	1315	10	Q9LIE8	arabidopsis
12	812	11.2	2187	11	Q9SPM0	zea mays (m
13	786.5	10.8	2112	5	P70670	mus musculu
14	784	10.8	1188	2	Q9VEL9	drosophila
15	715	9.8	763	10	Q41805	zea mays (m
16	709.5	9.8	6677	5	Q9XDH2	mycobacteri
17	707.5	9.7	555	10	Q9N435	caenorhabdi
18	703.5	9.7	4880	10	Q9FPQ6	chlamydomon
19	703.5	9.7	5085	11	Q9JLT1	rattus norv
					Q9JKS6	rattus norv
20	695	9.6	1151	13	O57580	gallus gall
21	684.5	9.4	1229	5	Q94185	caenorhabdi
22	674.5	9.3	2089	4	O14676	homo sapien
23	667	9.2	2284	5	Q9VPG1	drosophila
24	666.5	9.2	7962	4	Q10465	homo sapien
25	665	9.2	4833	11	Q90YX6	mus musculu
26	665	9.1	5038	11	Q90YX7	mus musculu
27	661.5	9.1	2768	5	Q9VC00	drosophila
28	653.5	9.0	1514	5	Q9GUW7	drosophila
29	653	9.0	3570	4	Q99552	homo sapien
30	643.5	8.8	990	13	Q91803	xenopus lae
31	636	8.7	3507	5	Q23587	caenorhabdi
32	636	8.7	6632	5	Q17362	caenorhabdi
33	633	8.7	489	10	Q41707	vigna ungui
34	632	8.7	761	10	Q92QI0	arabidopsis
35	631	8.6	2344	5	Q9N3Y8	caenorhabdi
36	629	8.6	1612	5	O01761	caenorhabdi
37	628.5	8.6	6642	5	Q9VYQ2	drosophila
38	627.5	8.6	971	5	Q9XVS4	caenorhabdi
39	622.5	8.6	839	2	Q9RX57	caenorhabdi
40	607.5	8.3	801	5	Q23635	caenorhabdi
41	607.5	8.3	924	12	Q99307	epstein-bar
42	605	8.3	379	5	Q27929	drosophila
43	600.5	8.3	1893	5	Q9NKC9	drosophila
44	598.5	8.2	409	10	Q9SBM1	volvox cart
45	591.5	8.1	2232	5	P91365	caenorhabdi

ALIGNMENTS

RESULT 1

ID   Q92954   PRELIMINARY;   PRT;   1404 AA.
AC   Q92954;
DT   01-FEB-1997 (TREMBLrel. 02, Created)
DT   01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   MEGAKARYOCYTE STIMULATING FACTOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA   FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA   Jacobs K., Turner K.;
RL   (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL   Mosher D.F. (eds.);
RL   Biology of vitronectins and their receptors., pp.45-52,
RL   Elsevier Science Publishers B.V. (1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Blood 78:279-279(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA   Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA   Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA   Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U70136; AAB09089.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR001212; Somatomedin_B.
DR   InterPro; IPR002400; GF_cysknot.
DR   Pfam; PF00045; Hemopexin; 2.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR00438; GFCYSKNOT.

```
DR   PRINTS;  PR00022;  SOMATOMEDINB.
DR   PROSITE; PS00024;  HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS05524;  SOMATOMEDIN_B; 2.
SMART; SM00120; HX; 2.
SMART; SM00201; SO; 2.
SQ   SEQUENCE   1404 AA;  151090 MW;  AABD7AD19B35F4F6 CRC64;

Query Match        99.6%;  Score 7245.5;  DB 4;  Length 1404;
Best Local Similarity   97.1%;  Pred. No. 0;
Matches 1363;  Conservative  0;  Mismatches  0;  Indels  41;  Gaps  1;

Qy    1 MAWKLPIYLLLLLSVFVIQQVSSQ------
        |||| ||||||||||||||||||||
Db    1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYINCQHYMECCPDF 60

Qy   26 ------ELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 79
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 120

Qy   80 PPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIW 139
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 PPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIW 180

Qy  140 KIKSKSNSAANRELQKKLKVKDNKKRNRTKKKPKPPVVDEAGSGLDNGDFKVTTPDIST 199
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 KIKSKSNSAANRELQKKLKVKDNKKRNRTKKKPKPPVVDEAGSGLDNGDFKVTTPDIST 240

Qy  200 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG 259
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG 300

Qy  260 KEKTTSAKETQSIEKKSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS 319
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 KEKTTSAKETQSIEKKSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS 360

Qy  320 TTPKEPTPTTIKSAPTTPKEPTTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEP 379
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTPKEPTPTTIKSAPTTPKEPTTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEP 420

Qy  380 APTTTKSAPTTPKEPTTTKGPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPK 439
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 APTTTKSAPTTPKEPTTTKKPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 480

Qy  440 EPAPTAPKKPAPTTPKEPAPTTTPKEPSPTTPKEPSPTTPKEPAPTTTKSAPTTKEPAPT 499
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 EPAPTAPKKPAPTTPKEPAPTTTPKEPSPTTPKEPSPTTPKEPAPTTTKSAPTTKEPAPT 540

Qy  500 TTKSAPTTPKEPSPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKKP 559
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TTKSAPTTPKEPSPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKKP 600

Qy  560 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEELAPTTPEEPTTPTT 619
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEELAPTTPEEPTTPTT 660

Qy  620 PEEPAPTTPKAAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 679
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 720

Qy  680 APTTPKPAPKELAPTTTKEPSTTSDKPAPTTPKGTAPTTPKEPAPTTPKG 739
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 APTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKG 780

Qy  740 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKEPAPTTPK 799
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPK 840

Qy  800 KPAPTTPETPPPPTTPSEVSTPTTTKEPTTTHKSPDESTPELSAEPTPKALENSPKEPGVPT 859
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 KPAPTTPETPPPPTTSEVSTPTTTKEPTTTHKSPDESTPELSAEPTPKALENSPKEPGVPT 900

Qy  860 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 919
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV 960

Qy  920 TSTTTQDTTPFKITTLKTTTLAPKVTTTTKKITTTELMNKPEETAPKDRATNSKATTPK 979
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 TSTTTQDTTPFKITTLKTTTLAPKVTTTTKKITTTTELMNKPEETAPKDRATNSKATTPK 1020

Qy  980 PQKPTKAPKKPFTSTKKKPKTMPRVRKPKTTPTPRKMTTSTMPELNPTSRIAEAMLQTTTRPN 1039
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 PQKPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTTSTMPELNPTSRIAEAMLQTTTRPN 1080

Qy 1040 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1099
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1140

Qy 1100 DETNICNCGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFT 1159
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 DETNICNCGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFT 1200

Qy 1160 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVY 1219
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESVY 1260

Qy 1220 FFKRGGSIQQYIYKQEPVQOKCPGRRPALNYPVVGEMTQVRRRFERAIGPSQTHTIRLQY 1279
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 FFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVVGEMTQVRRRFERAIGPSQTHTIRLQY 1320

Qy 1280 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV 1339
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKDQYYNIDV 1380

Qy 1340 PSRTARAITTRSGQTLSKVWYNCP 1363
        ||||||||||||||||||||||||
Db 1381 PSRTARAITTRSGQTLSKVWYNCP 1404
```

```
RESULT   2
Q9BX49
ID   Q9BX49;           PRELIMINARY;      PRT;   1404 AA.
AC   Q9BX49;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).
GN   BG174L6.2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wray P.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133553; CAC36090.1; -.
SQ   SEQUENCE   1404 AA;  782A1746B3FDEE5 CRC64;  782A1746B3FDEE5 CRC64;

Query Match        99.2%;  Score 7216.5;  DB 4;  Length 1404;
Best Local Similarity   96.8%;  Pred. No. 0;
Matches 1359;  Conservative  0;  Mismatches  4;  Indels  41;  Gaps  1;

Qy    1 MAWKTLPIYLLLLLSVFVIQQVSSQ------
        |||||||||||||||||||||||||
Db    1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYINCQHYMECCPDF 60

Qy   26 ------ELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 79
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP 120

Qy   80 PPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIW 139
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 PPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIR 180
```

```
Qy      140  KIKSSKNSAANRELQKKLKVKDNKKNRETKKKPTPKPPVVDEAGSGLDNGDEKVTTPDTST 199
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181  KIKSSKNSAANRELQKKLKVKDNKKNRETKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST 240

Qy      200  TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG 259
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241  TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG 300

Qy      260  KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKPEPTPTTKEPAS 319
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301  KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKPEPTPTTKEPAS 360

Qy      320  TTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTKEP 379
             ||||||||||||||||||||||||||||||||||||
Db      361  TTPKEPTPTTIKSAPTTPKEPAPTTPKEPAPTTKEP 420

Qy      380  APTTTKSAPTTPKEPAPTTPKEPAPTTPKKPAPTTPKEPTPTTPKEPAPTTKEPAPTTPK 439
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421  APTTTKSAPTTPKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTKEPAPTTPK 480

Qy      440  EPAPTAPKKPAPTTPKEPAPTTKEPAPTTPKEPSPTTTPKKEPSPTTPKEPAPTTTKKEPAPT 499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481  EPAPTAPKKPAPTTPKEPAPTTKEPAPTTTKEPSPTTPKEPAPTTTKKEPSPTTPKEPAPT 540

Qy      500  TTKSAPTTPKEPSPTTTKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTKKP 559
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541  TTKSAPTTPKEPSPTTTKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTKKP 600

Qy      560  APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEELAPTTPEEPTPTT 619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601  APTTKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPKEPAPTTPKEPAPTTKEP 660

Qy      620  PEEPAPTTPKAAAANTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKEPAPTTPKGTAPTTLKEP 679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661  PEEPAPTTPKAAAANTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEP 720

Qy      680  APTTPKKPAPKELAPTTTKKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKG 739
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721  APTTPKKPAPKELAPTTTKKEPTSTTCDKPAPTTPKGTAPTTPKEPAPTTPKG 780

Qy      740  TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKK 799
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781  TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKK 840

Qy      800  KPAPTTPEPPPPTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT 859
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841  KPAPTTPEPPPPTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT 900

Qy      860  TKTPAATKPEMTTTAKDKTTERDLRTTPETTTTNAAPKMTKETATTTTEKTTESKITATTTQV 919
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901  TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQV 960

Qy      920  TSTTTQDTTPFKITTLKTTTLLAPKVTTTTKKTITTTTEIMNKPEETAKPKDRAINSKAATTPK 979
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961  TSTTTQDTTPFKITTLKTTTLAPKVTTTTKKTITTTEIMNKPEETAKPKDRAINSKAATTPK 1020

Qy      980  PQKPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1039
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021  PQKPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080

Qy     1040  QTPNSKLVEVNPKSEDAGGAGEGTPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1099
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081  QTPNSKLVEVNPKSEDAGGAGEGTPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLS 1140

Qy     1100  DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRIEVWGIPSPIDTVFT 1159
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141  DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRIEVWGIPSPIDTVFT 1200

Qy     1160  RCNCEGKTFFFKDSQVWRFTNDIKDAGYPKPLFKGFGGLTCQIVAALSTAKYKNWPESVY 1219
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201  RCNCEGKTFFFKDSQVWRFTNDIKDAGYPKPLFKGFGGLTCQIVAALSTAKYKNWPESVY 1260

Qy     1220  FFKRGGSIQQVIYKQEPVQKPCGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1279
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261  FFKRGGSIQQVIYKQEPVQKPCGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1320

Qy     1280  SPARLAYQDKGVLHNEVKVSIIWRGLPNVVTSAISLPNIRKPDGVDYYAFSKDQYYNIDV 1339
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321  SPARLAYQDKGVLHNEVKVSIIWRGLPNVVTSAISLPNIRKPDGVDYYAFSKDQYYNIDV 1380

Qy     1340  PSRTARAITTRSGQTLSKVVWNCP 1363
             ||||||||||||||||||||||||
Db     1381  PSRTARAITTRSGQTLSKVVWNCP 1404
```

```
RESULT   3
ID      Q9JM99
        Q9JM99;              PRELIMINARY;        PRT;    1054 AA.
AC      Q9JM99;
DT      01-OCT-2000 (TrEMBLrel. 15, Created)
DT      01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT      01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE      MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE      PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN      PRG4.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Ikegawa S., Nakamura Y.;
RT      "a novel mouse gene highly homologous to a human gene encoding
RT      megakaryocyte stimulating factor precursor and cartilage superficial
RT      zone protein.";
RL      Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AB034730; BAA92310.1; -.
MGD;    MGI:1891344; Prg4.
DR      InterPro; IPR000585; Hemopexin.
DR      InterPro; IPR002965; P_rich_extensn.
DR      InterPro; IPR001212; Somatomedin_B.
Pfam;   PF01033; Somatomedin_B; 2.
DR      PRINTS; PR01217; PRICHEXTENSN.
DR      PRINTS; PR00085; HEMOPEXIN.
DR      PRINTS; PR00022; SOMATOMEDINB.
SMART;  SM00120; HX; 2.
SMART;  SM00201; SO; 2.
PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
PROSITE; PS00524; SOMATOMEDIN_B; 2.
SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;
```

Query Match   51.1%;
Best Local Similarity   54.5%;
Matches 772; Conservative 58; Mismatches 171; Indels 415; Gaps 26;

```
Qy        1  MAWKTLPIYLLLLLSVFVIQQVSSQ----------------------------------- 25
             |||||| |   |||| |::|||||||
Db        1  MGWKILPVCLSLILPVVLIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60

Qy       26  ------ELSCKGRCFESFERGRECEDCDAQCKRYDKCCPDYESFCAEVHNPTSPPSSKKAP 79
                   :|: :||::|   |:|::||||::|||||:|:|||||||:||||::||||||||
Db       61  KRVCSPELSCKGRCFESFEARGRECDCDSQCKQYGKCCADYDSFCEEVHNSTS-PSSKTAP 119

Qy       80  PPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSTIW 139
             |||||||||||| | ||| : ||| :|||||| ||||||:: ||||||||||||||
Db      120  TPAGASDTIKSTTKRSPKSPT-TRTIKVVESEELTEEHSDSENQE--SSSSSSSSSTIR 176

Qy      140  KIKSSKNSAANRELQKKLKVKDNKKNRETKKKPTPKPPVVDEAGSGLDNGDFKVT--TPDT 197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db      177  KIKSSKNS-ANRELQKNPNVKDNKKNTPKKKPNPEPAVDEAGSGLDNGEPKLTPPPDP 235

Qy      198  STTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTST 257
             |||:||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db      236  PTTPHSKVAISPKITTAAKPVTPKPSLAPNSETSKEASLASNKETTVETKETTATNKQSSA 295

Qy      258  DGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKPEPTPTTKEP 317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      296  DGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALITPKPEPTPTTKEP 317
```

Score 3716.5; DB 11; Length 1054;
Pred. No. 2.2e-233;
Pred. No. 2.2e-233; Length 415;

Db 1019 AFSKDQYYNIDVPTRTARAITTRSGQTLSKIWYNCP 1054

RESULT 4

O77765
ID   O77765;   PRELIMINARY;   PRT;   401 AA.
AC   O77765;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ARTICULAR CARTILAGE;
RA   Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA   Aydelotte M.B.;
RT   "Immunodetection and Partial cDNA Sequence of the Proteoglycan,
RT   Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT   Joints.";
RL   J. Orthop. Res.: 0:0-0(1998).
DR   EMBL; AF056218; AAD13404.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   Pfam; PF00045; hemopexin; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   SMART; SM00120; HX; 2.
FT   NON_TER   1
FT   NON_TER   401
SQ   SEQUENCE   401 AA;   44952 MW;   86147CC9AFBB73D7 CRC64;

Query Match   23.6%;   Score 1713.5;   DB 6;   Length 401;
Best Local Similarity 80.2%;   Pred. No. 9.3e-104;
Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;

Qy 966 PKDRATNSKATTPKPQKPTKAPKPTSTKKPKTMPRVRKPKTTPTPRK-MTSTMPELNPT 1024
            |||||||||||||||||| |:|||||||||| ||||| |||| ||
Db   5 PKGRATNSQVTTPKRPQKPTKAPKKPTSTKKPRT-PRVRKRPKTTPTPPKTTTSAMPEPPT 63

Qy 1025 SRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYL 1084
           : |||||||||| || |||||| ||||:|:|| || ||||||||| ||
Db   64 S-LPEAMLQTTTRPTPTPNSEIIDVNSENEDGDAAEGEKPHMIRPPVLTPIVIPGTEII 122

Qy 1085 PRVPNQGIIINPMLSDETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPSPARRI 1144
            : ||||||||||||||| |||||||||||||||||||||||||||||||| |||||
Db 123 VRGPSQGFGINPMFSDETNLCNGRPVDGLTTLRNGTLVAPRGHYFWMLTPFTPPPPRRI 182

Qy 1145 TEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVA 1204
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||:|||
Db 183 TEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKLISKGFGGLNGKIVA 242

Qy 1205 ALSTAKYKNWPESVYFKRGGSIQQYIYKQEPVQKCGPGRRPALNYPVYGEMTQVRRRFE 1264
           |||:||||:||| ||||||||||||:|||||| |||||| || |||||||||||||||
Db 243 ALSIAQYKSRPESVYFFKRGGSVQQTYKQEPTQKCTGRRPAINYSVYGETAQVRRRRFE 302

Qy 1265 RAIGPSQTHTIRIQYSPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGY 1324
           ||||||| ||||||||||||||||||:|||||||||||||||||||||||||||||||
Db 303 RAIGPSQVHTIRIHYTPVRVPYQDKGFLHNEVKVSTLWRGLPNVVSAISLPNIRKPDGY 362

Qy 1325 DYYAFSKDQYYNIDVPSRTARAITTRSGQTLSKVWVNCP 1363
           |||| |||||||||||||||||||||||||| | ||||
Db 363 DYYALSKDQYYNIDVPSRTARAITTRSGQTLSNTWVNCP 401

RESULT 5

Q9N4S7
ID   Q9N4S7;   PRELIMINARY;   PRT;   1079 AA.
AC   Q9N4S7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

Qy 1328 AFSKDQYYNIDVPSRTARAITTRSGQTLSKVWVNCP 1363
            ||||||||||||||||||||||||||||:||||||
Db 959 PFQTHTFRIHYSVPMRVSYQDKGFLHNEVKVSTMWRGFPNVVTSAIILPNIRKPDGYDY 1018

```
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y51B11A.1 PROTEIN.
GN   Y51B11A.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Cotton M.;
RT   "the sequence of C. elegans cosmid Y51B11A.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   1079 AA;  110532 MW;  8D8DE3824CF80CA1 CRC64;
```

```
Query Match      13.5%;
Best Local Similarity   29.6%;  Score 982.5;  DB 5;  Length 1079;
Matches  332;  Conservative  131;  Mismatches  481;  Pred. No. 7.7e-56;
                                                      Indels  177;  Gaps  44;
```

```
Qy    154 QKKLKVKDNKNKNRTKKKPTPKPPVVDEAGSGLDNGDFKV----------TTPDTSTTQ-HNKV 205
                                                          :||| :||:
Db     27 QKELEIADCRANQTPHTMLP-----------------STLISVDMETPSTLVLSSTPTSSSTPIKETT 77

Qy    206 STSPKITTAKPINPRPSLPPNSDfSKEtSLIVNKEtT-VEtKEtTTTNKQtSIDGKEKTt 264
           :||:: ||                         :|||  :|||: ||::
Db     78 TTTAPETTSTEP-------------PSSSTTPVQTTTTTAPETTSTEAPSSSTTPVQTTT 124

Qy    265 SAKEtQSIERKTSAKQLLAPTSKVLAKPTPKAETT-TKGP-ALTTPKEPTPTTPKEPASTTP 322
                                          |||  ||| :||||| | || |||||||
Db    125 TAPETTSTEPPSSS----TSPVQTTTTTAPETTSTEAPSSSTTPVQTTTTTAPETTSTEP 180

Qy    323 KEPPTTIKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 382
          | || || ||||||||| ||||| ||: | ||||| |:: |
Db    181 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPE---TTISTEP-PSSSTTPVQT 236

Qy    383 TTTKSAPTTPKEPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPA 442
          ||| ||||||||||||||| |||||| ||||| | ||||| ||||
Db    237 TTTTAPETTSTESPSSSTTPVQTTTTTAPETTSTEP-PSSSTTPVQTTTTTAPETTSTEP 295

Qy    443 PTAPKKPRAPTTPKEPAPTTPKEP-----------APTTTKEPSPTTPKEPAPTTTKS 488
          | ||| || |||||||||||||              ||||||| |||||||||||
Db    296 PSSSTTPVQTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTT 355

Qy    489 APTTTK--EPAPTTTKSAPTTPKEPSPTTPKEP-APTTPKEPAPTTPKKPAPKELAPTTPKGPTTSTPTTSDKP- 545
          || |||   | ||||| ||||||| |||||||  |||||||||||    |||   |||||||||: |
Db    356 APETTRTEPPSSSTTPVQNTTTTAPETTTSTEPPSSSTTPVQTTTTT----APETTSTEPP 411

Qy    546 TTPKEPAPTTTKKPAPTAPKEP---APTTPKEPAPTTPKKLTPTTPEKLAPTTPEKPAPTT 603
          :|| ||||||  |  | || ||   |||||||||||| :: | || || ||||||||||||
Db    412 SSSTTPVQTTTTTAPETTSTEEPPSSSTTPVQTTTTTAPETTSTEP----PSSSTTPVQTT 467

Qy    604 PEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEP--APTTTPKEPAPTTTKEPAPTTPK 661
          | | ||||| || ||||| |||||||    || || |||||| ||||| ||||| ||||
Db    468 TTTTAPETTSTEP-PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPE-TTSTES 525

Qy    662 ETPAPTTPKGTAPTTLKEPAPTTKPAPKELAPTTTKEPTSTTSDKP-APTTPKGTAPTT 720
          : || || | |||| ||||||| ||||| |||||||| |:: |  || | | |||| 
Db    526 PSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTT 585

Qy    721 PKEPAPTTPKEP--APTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTTTPVQTTTT 777
          | | ||||||||  ||||| | |||| ||||||||  ||||| |||||: |||||||||||
Db    586 APE---TTISTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPS 642

Qy    778 APTTPKETAPTTTPKEPAPTTPKK-----PAPTTPETPPPTTS------EVSTPTTTKE 824
          ||||| |  || ||||||||||||     ||||||| || |        |:||||||||
Db    643 SSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAP 702

Qy    825 PTTIHKSPDESTPELSAEPT--PKALENSPKEPGVPTTKPAATKPEMTTTAKDKTTTERD 882
          |||  ||| |  | | |||   |||| ||| |||| |||| | |||:|||||||||||||
Db    703 ETTTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTP 762

Qy    883 LRTT---PETTTAAPKMTKETATTTEKTTTESKIFATTTQVTSTTTTQDTTPFKITTLKTT 938
          :|||   |||| || |:|:| |||| ||||| || |||||| |||||| |:||| ||||||
Db    763 VQTTTTTAPETTSTEPPSSSNTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTE 820

Qy    939 TLAPKVTTTKKITTTTTEIMNKPEETA-KPKDRATNSKATTPKPKQKPTKAPKKPISTKKPK 997
          : || |||||| ||:: | || ||||  |||  ||| |||||| |||| |||| || |||||
Db    821 STEPPSSSSNTPVQTTT--TTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSTTPV 878

Qy    998 TMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTIRPNQTPNSKLVEVNPKSEDAG 1057
          | || |||||||| |:||||||||| | | ||||||||:||||||||| |:|:||||| |
Db    879 QTTTTTAPETTSTEPPSSSTTPVQTTTTTAPET---TSTEP--PSSSTTPVQTTTTTTAP 932

Qy   1058 GAEG-ETPHMLLRPHVFMPEVTP---DMDYLPRV-------------PNQGIIINPMLSDETN 1103
          :| | ||| ||||| |:||| ||   ||:|||||              |||:|||| |: ||
Db    933 ETTTSTEPPSSSTTPVQTTTTTVPCDCSLSYIDRVVYPFTTEWWENKRDIIIQSYDSPRRT 992

Qy   1104 ---------------ICNGKPVDGLTLRNGTLVAFRGHYFWMLS--------PFSPPS 1139
                         ||:|||||||||||||| |||||||||||||        |||||:
Db    993 AFSFVSTFDIGTCTRAILTCIYTSEGGISNL-NATL---------FIGLSDGSSIDLPF---- 1040

Qy   1140 PARRITEVWGIPSPIDTVFFTRCNCEGKTFFFKDSQYWRFTN 1180
          | |:| |||||| ||| :| ||:|||||:||:||| ||| |
Db   1041 ----YNVPIGLE-IWPEINCEGNK--------WSTYNN 1064
```

```
RESULT   6
ID   Q9I7S1        PRELIMINARY;      PRT;   1049 AA.
AC   Q9I7S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
```

Query Match
Best Local Similarity 13.0%; Score 946.5; DB 5; Length 1049;
Matches 376; Conservative 34.4%; Pred. No. 1.6e-53; Mismatches 480; Indels 169; Gaps 56;

```
Qy   65 EVHNPTSPPPSSKKAPPSGASQTIKSTTKRSPKPNKKKTKKVIESEEITEEHSVSENQE 124
Db   33 EIGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPT---EGTTAKPTTLKPTE 86

Qy  125 SSSSSSSSSTIWIKIKSSKNSAANRELQKKLKLVKDNKKNRTKKKPTPKPPVVDEAGSG 184
Db   87 GTTAKPTT------KPTTLKPTE--GTTAKPTTLKPT----EGTS 128

Qy  185 LDNGDFKVTTPDTS--TTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETT 242
Db  129 AKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKP---TTLKPTEGTSAKPTT 185

Qy  243 VETKETTTTTNKQT--STDGKEKTTSAKETT--OSIEKTTSAK--DLAPTSKVLAKPTT---PK 293
Db  186 LKPTEGTSAKPTTLKPTEG---TSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPT 241

Qy  294 AETTTKGPALTTPKEPPTTPKEPAPTTPKEPA---PTTTKS 347
Db  242 EGTTAK-----PTTLKPTEGTTAKPTTLNPTEGTSAKPTTLKPTEGTTAKPTTLNPTEGTS 297

Qy  348 APTTPKEPAPTTPKEPA---PTTPKEPAPTTPKEPAPTTKSAPTTKEPAPTTPKKPAP 404
Db  298 AKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLK---PTEGTSAKP 352

Qy  405 TTPKEPAPTTPKEPT---PTTPKEPAPTTPKEPA---PTAPKKPAPTTPKEPAP 459
Db  353 TTLKPTDGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEG 412

Qy  460 TTTPKEPAPTTTKEPSPTTPKEPAPTTPKSAPTTTKEPA---PTTTKSAPTTTKPEPSPTTT 516
Db  413 TTTAK---PTTLKPTEGTTAK---PTTLKPTEGTSAKPTTLKPTEGTT 466

Qy  517 KEPAPTTPKE---PAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKKPAPTTPKE---PAPT 570
Db  467 AKPTTLKPTEGTTAKPTTLKPTKGTSAKPTTLK-PTEGTTAKPTTLKPTEGTSAKPTT 525

Qy  571 TPKETAPTT---PKKLTTPTTPPEKLAPTT--TPEEPTTPTTPEEPA 624
Db  526 TLKPTKGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGT 585

Qy  625 PTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKPEPAPTTP 684

Db  586 TAKPTTLKPTEGTSAQPTTLK---PTEGTSAQSTTLKPTEGTTAK---PTTLK---PTEG 636

Qy  685 KKPAPKELAPT------TTTKPTSTTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTP 737
Db  637 TSAQPTTLKPTEGTTAKPTTLKPTEGTSAK--PTTLKPTEGTTAK---PTTLKPTDGTTA 691

Qy  738 KGTA-----------PTTLKEPAPTTPK---KPAPKELAPTTTKGPTSTTTSDKPAPTTPK 783
Db  692 KPTTLNPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPT 751

Qy  784 ETAPTTPKEPAPTT--PKKPAPTTPTEPPPTTPTTISEVSTPPTTTK------EPTTIHKSPDE 834
Db  752 EGTRTTAK---PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTL--KPTE 806

Qy  835 STPELSAQPTTP-KALENSPKEPGV--PTTKT---PAATKPEMTTTAKDKTTERDLRTTPE 888
Db  807 RT---SAQPTTLKPTEGTSAKPTTLKPTEGTSAQPTTLKPTEGTTAKPTT---LKPT-E 858

Qy  889 TTTAAPKMTKETATTTEKTTESKITTATTTQVTSTTQDTTPFKITTLKTTTTAPKVTTTK 948
Db  859 GTSAKPTTLKPTEGTTAKPTT----TLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTA 914

Qy  949 K---TITTTEIMNKPEETAKPKDRATTNSKRATTTRPKQKPTKAPK---KPT--STKKPKTMPR 1001
Db  915 KPTTLKPTEGTSAKPTTLKPTE-GTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKP 973

Qy 1002 VRKPKTTPT---PRKMTSTMP-ELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAG 1057
Db  974 TEGPSAKPTTLKPTERTSAQ-PTTLKPTEGTTAKPTTLKPTKGTSG 1031

Qy 1058 GAEG-ETPHMLLRP 1070
Db 1032 QANNFETKRKERRP 1045
```

RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003495; AAG22353.1; -.
DR FlyBase; FBgn0030561; CG5228.
SQ SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;

RESULT 7
O76894

ID   O76894   PRELIMINARY;   PRT;   1795 AA.
AC   O76894;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:56G7.1 PROTEIN.
GN   EG:56G7.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nixon K.A., Nusskern D.R., Pacleb M., Palazzolo M.,
RA   Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003421; AAF45644.1; -.
DR   EMBL; AL031028; CAA19845.2; -.
DR   FlyBase; FBgn0025390; EG:56G7.1.
DR   InterPro; IPR0025557; Chitin_binding.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF01607; Chitin_bind_2; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00494; ChtBD2; 2.
SQ   SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match   12.3%;   Score 892; DB 5; Length 1795;
Best Local Similarity   26.5%;   Pred. No. 9.8e-50;
Matches 319; Conservative 154; Mismatches 520; Indels 210; Gaps 38;

Qy   21 QVSGQELSCKG-----------RCFE--SFERGRECDCDAQCKKYDK-------CCPD 58
                                    ||:|| |||
Db   197 QCPSTEISDSGSYIPQNCELKFPECAEECTFRSPTDCALYYTCRLQESGTYLQCTRFKCPG 256

Qy   59 YESFCAE-----------------VHNPTSP---PSSKKAPP-------PSGASQ-- 86
               || ||||| ||| |:||
Db   257 SNSFDLERLCRPRSEVDCFDFVPGPVQVPYAPQPYPYPAAPLYEEDYDIGAREQQ 316

Qy   87 -TIKSTTKRSPKPPNKKKTKKVIESEEIIEEHSVSENOQESSSSSSSSSSSSSSTTWKIKSSK 145
            :|| |||:   :
Db   317 PALKSEKLQVAAEGFEKPSLNVVVLQTTTLEPSTAVHKYPAYPSYPSYEYSS--HHRGKE 374

Qy   146 NSAANRELQK------KLKVKDN------ 167
           :|:| ||:| :|
Db   375 RAAENLELEKEGVPRKLKLSENIVIQOPETPATATAATTREPLNDINKYQVKRYTYGTDKNDV 434

Qy   168 KKKPTPKPPVVDEAGSGLDNGDFKVTPDTSTTQHNKVSTSPKITTAKPINPRSLPPNS 227
              |:|| ||::|| : |||| :|| |||:| |:||
Db   435 TEAPEIKSPL----KGLHLSENIVILPETTTT--TTTTTKPVVLCPTISPPDTPFKPS 487

Qy   228 DT--------------SKEISLTVNKEITVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTS 276
                             :|| ||||| :||||| ||::||
Db   488 TTTAVTKSTTKPKISSTEQHSTTTAKTTTTKRPTTVTEKTSSAATEKPRTTVVTTTQKRSTT 547

Qy   277 AKDLAPTPSKVLAKPTPKAETTTKGPALTTMTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTT 336
             ||:||| || | ||| :| | || ||||||||||||||||||||:||||
Db   548 THNTSPDPKTTIRSTTLSPKTTTTPSTTTTPTTPSTT----TTPSTTTPST----TT 596

Qy   337 PKEPAPTTTKSAPTTPKEPAPTTTK-EPAPTTTKPKEPAPTTTKEPAPTTTKSAPTTPFKEPA 395
           |:||: | || :|:| || |||| || |||||||| || |||| || || :|: ||
Db   597 PSTTTPSTTTVKVSTTRPRTTSQKTTTASTTTKK-----TTTSPKTTKTTDIPTSTTSKL 652

Qy   396 PTTPKKPKAPTTPKEPTPTTPKEPPTTTTPKEPAPTTPKEPASTTTSPKPTSSTGKPTTPK 455
           |||| :|:||||||| ::|| || |||||||||||||||:|| |: |: |||:
Db   653 STTTQKTTTTHKFTAATTTSTEKPKTTTTEKTSTVSTTTTKKSTESSPKPTSSTGKPTTPK 712

Qy   456 EPAPTTTKSAPTTKEPSPTTTKEPAPTTTTKSAPTTTPKEPAPTTTKSAPTTPKEPSPTT 515
           ||| |||||||||||:||||||||| | | ||||||| || |||||||||||:|||
Db   713 PSTRTTPTTTKVTTTTQITTTTPLRSSTETITSTQPPTTTTQPPTTTTTLTVTPKTSTTTTT 772

Qy   516 TKEPAPTTTKEPAPTTPK--------KPAPTTPKEPAP--------TTPKEPAPT 554
           :|| ||||||| ||||| |||||||||| ||||||
Db   773 TTEKPITTSSPKPTTTTTSTAPNTTKVAITTQKETTTPTQSTSTTIFTRKTTTNPEPT 832

Qy   555 TTTKKPAPTTAPKEPAPTTTPK-ETAPTTTPKKLTPTTPEKLAPTTPEKP-------APTTPE 605
           || | ||| |||||| || | |||||||||||||||:|||||||| ||||||
Db   833 STEKPITSTTTKPSTTTKTSTVASSTEKTTISSPKPTTTEKSTENPTTNSVKTSALTSST 892

Qy   606 ELAPTTPEEPTTPTTPEEPAPTTTPKAAAPNTMKEPAPTTTPKEPAPT--TPKEPAPTTPKET 663
           ||| ||| ||: | || || : :|| || || |||||||| |||| ||:|:
Db   893 QRATSTTSEPTKTT-QNITTTTPKPTTLKTSTQEATTSTQOKVSTVTITTKKATESSPLTT 951

Qy   664 APTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTPKEPT--STTSDKPAPTTPKGTAPTTP 721
            || | |||| ::|| || :|||||| |||||| |:| |||| | |||||
Db   952 LSTEEPNTTPKPLRTTTPTTPTTVSVTATVTRITPTTTSESSTETTTSTOQKPKSTTTPTSTTRTTP 1011

Qy   722 KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTTPKKPAPKELAP-TTTKGPTSTTTSDKPAPT 780
           || ||||||| ||||| |||| |||| |||||||||:||| |||||||::| ||||
Db   1012 KVTIVIVSTTQNPTTTSKTSTVTI-----TTP-NPSPSTQRPTTTTTTTRQPTSTVASTTSIG 1065

Qy   781 TPKETAPTTTPKEPAPTTPKKPAPTTPEPPPTTISEVSTPITT-TKE---------- 824
           | | ||||||||| ||| | | |||||||:|| | ||| |||
Db   1066 TTRIPTTTNPQNSTSSTDLTTVTRPPCDPDDSTSDKNTNTACTOELQQVNLELQSPQKQ 1125

Qy   825 ----PTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTE 880
                :||| || || ||| |||:|| ||:||| || :
Db   1126 EQFTHTRTHIALTGSRNTLGGQEVPDTVMDDAPSSAEAESGQATTAKAPTMSTLAAANLLQ 1185

Qy   881 R---DLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPPKITTLKT 937
                :||| ||| ||||:|| |||| | ||:|||:|||||
Db   1186 KLFHIISTTPPSREHAP--TQRSSQPSSQOSR-GVTIAQMARHNILATSKPPIAHSLRL 1242

Qy   938 T----TLAPKVTTTKKTITTTELMNKPEETAKPDKRATNSKAPTTPKPRQKPTKA-PKKPTS 992
                ||| | ||| | | || | | |||||| :|| ||||:| || | ||||||
Db   1243 SIQOLASTQKRSIPPKTLVTHNTTKEPED-SEYVDSEITSEQTTDEDNEVLDKTQOPRAMSS 1301

Qy   993 TKKPKTMPRVRKPKTTPPPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPK 1052
           | | |||||||||| |||| :|| | |||| || |||:| ||| ||:||
Db   1302 TTVAAVLPAVPSTTTEREPQK-TSSSP--SPT----KATSSTTTQPIETTTIGDLEVDSSG 1354

Qy   1053 SED 1055
           |||
Db   1355 SSD 1357

RESULT   8
096449
ID   096449;   PRELIMINARY;   PRT;   1489 AA.
AC   096449;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Goernhardt B.;
RC   STRAIN=RACE 1-11;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.

```
DR   EMBL; AF061185; AAC72308.1;
SQ   SEQUENCE   1489 AA;  164037 MW;  764CC79D1C2F5163 CRC64;

Query Match           11.9%;   Score 863.5;  DB 10;   Length 1489;
Best Local Similarity 31.7%;   Pred. No. 5.7e-48;
Matches 368;  Conservative 55;  Mismatches 509;  Indels 229;  Gaps 51;

Qy  104  TKKVIESEITEEHSVSENQESSSSSSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK  163
              ||: |: |:    :         :||    ||
Db  282  TPTVGYSTETEGQOHVTGGYEPSDEIEAPTEGTTY--VPREETTAAPSE----DTTYAP      334

Qy  164  KNRIKKKPFPKPPVVDEAGSGLDNGDFKYTTPDITT----QHNKYSTSPKITTAKPI        217
                      |||       |  : |:|      ||| :|
Db  335  REVTPYAPIEKPYDVEETTYVTEESTYAPIKSEINAPTERMHYAHIEKPCDIEVTMYAPT     394

Qy  218  NPRPSLP----------PNSDT---SKETSLTVNKETTVETKETTT----TNKQT        255
                              :|| ||         |:|: | | :|      |:||
Db  395  EETTYAPTEETTYAPTEEMYAPIEETTYAPTEETPYEPTEETTYAPTEETTYAPTEKTT      454

Qy  256  STDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTIKGPALTTPKEPTPTPK    315
              | |||  |:|        ||          ||      || |  |||
Db  455  YAPTEETTYAPTEETPYEPTEETTYAPTEEETTYAST...EETTYAPTEETTY           511

Qy  316  EPASTTPKEPTPTTIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT--TTKEPAPTTPKEPA  372
              |||  ||  ||| |||      |||       |||        |||
Db  512  APAEETPYEPTEET-TYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPT           570

Qy  373  PTTTKEPAPTTKSAPT------------TPKEPAPTTPKPAPTTPKEPAPTTPKEPTPTT    422
              ||           ::|      ||   | |    ||     | ||
Db  571  EETTYAPTEETT-YAPTEETMYAPIEETTYAPTEETPYEPTEETTYAPTEET          629

Qy  423  PKEP--------------TTKEPAPTTPKEP-----APTAPKKPAPTTPKEPAP        459
              |||               ||        ||     | | | | ||
Db  630  TYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAP      689

Qy  460  TTPKEPAPT--TTKEPSPTTPKEP-----APT-TTKSAPT-----------TTKEP----  496
              ||  | |  |||     |  ||     |||  |||            |||
Db  690  TEETTYAPTEETTYAPAEETPYEPTEETTYAPTEEMYAPIEETTYGPTEET           749

Qy  497  --APT-TTKSAPT--TPKEPSPTTTKEP-----APTTPKEPAPTTPKPAPTTPKEPAP     545
                |||  |||       |  :   ||     |||     |||     || ||
Db  690  TEETTYAPTEEETTYAPTEETPYEPTEETTYAPTEETTYAPTEETTYGPTEET           749

Qy  546  T--TPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPKPAPTT    603
              |   | |||  || |:  :| :| |  |     |  | ||   ||
Db  750  TYAPTEATTYAPTEETPYAPTEETTYAPTEETTYGETTYAPTEETTYAPTEETTYAP        809

Qy  604  PEELAPT--TPEEPTPTTPEEP---------APT--------------TPKAAAPNTPKEP  645
              |||   | |||  ||          |||              || | | |||
Db  810  TEETPYEPTEETTYAPTEEPTYEPTEET---TVTPTEETTYAPTEETTYAPTEETPY       867

Qy  696  -TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP-----APTTPKGTAPT----   743
              ||              | |:||     |||      |||
Db  928  EPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETPYEPTEETTYAPTE            987

Qy  744  ------------TLKEPAPTTPKKP------APKE---LAPT--TTKGPTSTTSDKPAP   779
                          |  |||  ||       | |     | |    || |
Db  988  EETTYAPTEETMYAPIEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTY          1047

Qy  780  TTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTSEVSTPTTTKEPT---TIHKSPDRSTP   837
              |||  |||  |    |                ||   |||
Db 1108  ETTYAPAEETPYEPTEETTYAPTEETTYAPIEETTYGPTEETTY-APTEATT         1166

Qy  838  ELSAEPTPKA-LENSPKEPG------VPTTKTPAATKPEMTTYPAKDKITTERDLRTTP--  887
              |:|  ||      |  ||              :|       ||  |     |
Db 1167  YAPTEETPYAPTEETTYEPTGGETTYAPTEETTYAPTEETTYAPTEETTYEPT         1226

Qy  888  EETTYAAPKMKETAFTTEKTTESKITATTTQVTSTTIQDTTT---------PFKITT--  934
              |||                      |      ||         |||
Db 1167  YAPTEETPYAPTEETTYEPTGGETTYAPTEETTYAPTEETTYAPTEETTYEPT         1226
```

RESULT   9

ID   Q20007   PRELIMINARY;   PRT;   1274 AA.
AC   Q20007;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE   COSMID F35A5.
GN   F35A5.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Leimbach D.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U46675; AAB52641.1; -.
SQ   SEQUENCE   1274 AA; 07F6B0D292C1799F CRC64;

```
Db 1227  EETTYAP-------TEETTYEPTEETTYAPTEETTYAPTEETTYAPIDETTYGP        1281
              |||  |        |||     |||   ||| |||
Qy   935  LKTTTLLAPKVTTTKKITTTEIMNKPEETAXPKDRAT--NSKATTPKP-OKPTKAPKKPT   991

Db 1282  TEETTYAPTEATTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPMEET  1341
              |||  |||  | |||     |||
Qy   992  ----------STKKP-KTMPRVRKPKTTPTPPRKMTSTMPELNPTSRIAEAMLQTTTRP   1038

Db 1342  PYEPAEESTSTVSTEKPICNTEEFTDEPTDEPTDLPTDGPSTP                   1399
              ||| |||     :|   ||     ||           |||
Qy  1039  --NQTPNSKLVEVNPKSEDAG                                         1057

Db 1400  CDNQGINGIGVENKVRYNNAG                                         1420
```

Query Match   11.6%;   Score 846; DB 5;   Length 1274;
Best Local Similarity 27.6%;   Pred. No. 6.7e-47;
Matches 313; Conservative 119; Mismatches 455; Indels 248; Gaps 60;

```
Qy   73  PSSKKAPPSGASQTIKSTTKRSP----KPP---NKKKTKVIESEITEEHSVSENQES     125
              |||            |           ||        ::| |||
Db  185  PSPKKAAPSKEHDPIVPPTIKNPAKKWKPWEDDEVPTEEIKEPEEPATRKVPALKKKEP   244
```

```
Qy     126  SSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK-KNRTKK-KP------TP---- 173
            ::|:| .|| .|             |.|||      || ||| |:
Db     245  STSVKPVSDPSPPIKKV----------------PVKKEPEVPPTPIKNPTKKWKPPWEDETPVEEVK 294

Qy     174  KPPVVDEAGSGLDNGD---------FKVTTPDTSTTQHNKVSTSPKITTAKPI-NPRSLPPNSD 228
            ||||:                   || || ||| ||:||||     :||||
Db     295  EEPVPEKKAPVLKKKDPAPAAARDPSPSKAAPKKVEPSSPVVPPTPVKNPVKYKPPWE 354

Qy     229  TSKETSLTVNKETTVETKETTITNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLA 288
            :||||        |       .|| ||:|        ||:|  ||  ||||:| |||:
Db     355  VDDEPAEEVKKPSAPEKKTPVLKRKEPEPSSTTPSSDPSPKKAAPAVKPDSSPKKATPL 414

Qy     289  KPTPKAE----------TTTKGPA---- 323
            |||:|            |||:
Db     415  QADPKAQEVPPTPVVKNPVKKYKPPWEVDDEDPVEEVKQPEAPAKKTPVLKRKEPAAKDTA 474

Qy     324  EP------TP-TTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPA---- 364
            :|:     |||  ||| :|||  |||||    ||||||
Db     475  KPATSKPETEPKKDPVKPRDSSPKKVAAKPDSAQAPA-TPVKNPVKWRPPWEDDETPA 533

Qy     365  ------PTTPKEPAPTTTKEPAPT-----TTKSAPTTPKEP---- 407
            |||||  :|          |||||||
Db     534  DDVSKPTDAKKTPSLAKKDPAPAKESLKPKADTKAPAKPRDPSPRKVVAAPTAPEKKTPVLA 593

Qy     408  KE-----PAPTTPKEPPTTPKEPAPTTPKEPAPTTPK-EPAPTAPKKPAP----- 456
            :|:     |||          |||:  ||   |||  ||||||
Db     594  KKEPAGPADSKTKEPEKSKPRDPSPRKKVAAPKPVEKTEVAPAAVKKPEPISKPKDTAPKK 653

Qy     457  PAPTTPKEPAPTTTKEP------SPTTPKEPAPTTTKSAPTTTKEPAPTTTKEP-KEP 511
            |||||||        |||||     ::|           |||||    ||:
Db     654  AEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPEPEKK-TPVLAKKAPTKPDSEA 711

Qy     512  SPTTTKEPAPTTPK--EPAPTTPKKPAP--TTPKEPAPTTPKEPAPTTTKKPAPTA-PKE 566
            |||::  ||||   ||| |                 :|||| ||
Db     712  AADPVSGPSSKDPKLAKKAAPVKPRDPSPMKAVPIKPAPKT--EVPPAAVVKKPEPVAKSRD 769

Qy     567  PAPTTPKETAPTTTPKKLTPTTP----EKLAPTTPKRAPTTPEELAPTTPEEP-----T 616
            |||||||||        :|||           ||| |  ||| ||||| ||||
Db     770  PSPKKAK-AEPNSP--VVPPTPVKNPVKWKPPWEDDDAPAEPVNVPEPEKKTPVLAKKT 826

Qy     617  PTTPEEPAPTTTPKAAAANTPKEPAPTTTPKEPAPTT-PKEPAPTTPKETAPTTPKGTAPTT 675
            ||| :|||  | |||         |||||||          |||||   |||||||
Db     827  PVKPRDPSPRKKAVPAPAKPSKTDAPPVSVKKPEPVSKPEPSPKKAEPNSPVVP-----PTP 882

Qy     676  LKEPAPTTPKKPAPKELAPTT--TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAP 734
            :|||||||            || | |         ::||||||  |||||||||
Db     883  VKNPVKKW-KPPWEDDDEPTEEVKKPSE--PEKKTPVLAK-KEPEKPKD-APKVAAKPRD 937

Qy     735  TTPKGTAPTTLKEPAPT-----TPKKPAP-----KELAPTTKGPTS-----TTSDK 776
            |||||| :||||              |   |||  |||||||||
Db     938  PSPKKAVPE---KEPAKVAAKPRDLSPKKAIPIPANTQEAPPTPVKNPVKKWKPPWEDDDE 995

Qy     777  P--------APTTPKET---------PTTPIAKPPIATPPIT------TPPVATPPITT 825
            |         |||||||         |||
Db     996  PAEPVSAPEPEKKTPVLAKKAPAKPRDP---SPKKAAPVAAK--PDPKIPEV-PPTPVKNP 1050

Qy     826  TTIHKSP-----DESTPELSA-EP----TPKALENSPKEPGV----- 864
            |||:|           ||  ||    |||| :||| ||
Db     1051 VKKWKPPWEDDDEPSEPVSAPEPEKKTPVLAKKAPTKPATKPDSEAAADPVSGPTSKDPK 1110

Qy     865  AT-----KPEMTTAKDKTERDLRTTPETT--TAAPKMTKETATTTEKTTESKITPATT 916
            ||      |||           ||||
Db     1111 LSKKAPVEKPKPTTDPKDDKLKPSPAKKPEKAPEPAAPKKWKPVWDDDPDEPEADTVPA 1170

Qy     917  TQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETTAKPKDRAATNSKAT 976
            |       |||      |         |
Db     1171 PSRKKPDIEDPADPLG------GPRTKDPK------LNKKAPAEKPTEK------ 1206

Qy     977  TPKPQKPTKAPKKPTSTKKPKTMPRVRKPK------TTPPRKMTSTMP 1019
            || :      :|  |       | |:         :| :| ||
Db     1207 -PKPKEVSKEPKPKTEPKKP-AAPKKWKPPWEDDDEPEADTMPARKKPDTEDP 1259
```

```

Query Match             11.2%;     Score 817;   DB 10;   Length 1480;
Best Local Similarity   28.4%;     Pred. No. 5.9e-45;
Matches  343;  Conservative  63;  Mismatches 563;  Indels 240;  Gaps 54;

Qy       4  KTLPIYLILLLSVFVIQQVSSQELSCKGRCFESFERGRRECDCDAQCKRYDKCCPDYESFC 63
            :|||||.| |||||||| |||          |       || | |    ||||
Db       6  RNLSFLILLLNFFAATY---------------------ARECSPPKPSPKPHK--PPKHSVV 45

Qy      64  AEVHNPTSPPSSKKAPPSGASGQTIKSTTKR----------SPKPEN------KKKTKKVIESEEI 113
            |              |   || | ||:|               |         |||||:::
Db      46  -----PPKEPAAVKPHPHPK--PPTIKPPPPKRHPHKPPTVKPHPHKPPTKPHPHKPP 98

Qy      114 TEEHSVSENQESSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVK--DNKRNRTKKKP 171
            ||| || | |      || |  ||                       |   | :|
Db      99  TKPH----------------PHPKPPTIKPPPHPK----PRPHPKPPNVKPHPHPKPHP 141

Qy      172 TPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPPNSDTSK 231
            |:||              | |:|                :| | || ||
Db      142 HPKPP----------------TKHH----PHPKPPTIKPPPKPPSVKPPPSTPK 175

Qy      232 EISLITVNKETTVETKETTITNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV--LAK 289
            |||                                             ::||
Db      176 --------------------PPTTNPPPSTPQPPPIHKPPPCTPPPVASPPMATPPTQMPPIAT 219

Qy      290 PTPPKAETTTKGPALTTPKEPTP--TTP--KEPASTTPKEPPTTTIKSAPTTPKEPAPTTT 345
            |  | | :|||||| ||| | |    ||  ||||
Db      220 P-PIAKSPVATPPIATPPITTPPVATPPITTPPIANPPIIMPPIATPPVAAPPIT 278

Qy      346 -----KSAPTTPKEPAPTTT----------------KEPAPTTPKEPAPTTTKEPA---- 380
            |||||||        |||||||||
Db      279 NPPISKPPVTTP-----PTTTPPIAKPPIATPPISTPPAATPPITTLPPAKPPVAI 334

Qy      381 -PTTTIKSAPTTP-KEPAPTTPKKPAP----TTPKEPAPTTPKEPTPTTTKEPAPTTT 436
            |  ||||||||| ||||||||                 |||||||   || || ||
Db      335 SPIVT--PPVTPIAQPPVATPTTATPPVATPPIATPPTSKSPISTPPISESPVATPPTAT 392

Qy      437 TP-KEPAPTIAPKKPAPTTPKEPAPTTPKEPAPTTT-----KEPSPTTPKEPAPTTTKSAPT 491
            || |  || || |  |   ||||||| |||||||||     ||||  ||||||||| |||
Db      393 SPIKTPPPAKPPVATPPIAKSPIATPPTATPPIEKPPVATPPTTPPTAT---PP 449
```

```
RESULT  11
Q9SPM0
ID   Q9SPM0          PRELIMINARY;   PRT;   1315 AA.
AC   Q9SPM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN.
GN   PEX2.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=POLLEN;
RA   Hohorst D., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA   Stratford S., Bedinger P.A.;
RT   "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159297; AAD55980.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR001998; Xylose_isom.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR   SMART; SM00370; LRR; 3.
DR   PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
SQ   SEQUENCE 1315 AA;  134401 MW;  64C97A2A01F0936F CRC64;

Query Match          11.2%;   Score 816;   DB 10;   Length 1315;
Best Local Similarity  23.5%;   Pred. No. 6.1e-45;
Matches  265;   Conservative 106;   Mismatches 486;   Indels 270;   Gaps 35;
```

```
Qy    32  RCFESFERGRECDCDAQCKKYDKCCPDYESFGCAEVHNPTSPPSSK-KAPPPSGASQCTIKS   90
Db   390  QCAPVLARPVEVDCS----------------------KHVCAGYPTPGGPSSSVPGKPPSSVPGK----  432

Qy    91  TTKRSPKPPNKKKKTKKVIESEEITEEHSVSENQOESSSSSSSSSSSSSSSSTIWKIKSSSKNSAAN  150
Db   433  --PAAPAPMPTHIPPPDVSPEPLPEPSPV-----------------------PAPAPMRMPTLRSPPAD  476

Qy   151  RELQKKLKVKDNKKNRTKKKPPTKKPPVVDEAGSGLDNGDFKVTTPDYTSTIQHNKVSTSPK  210
Db   477  EYI--------------------------PTPPVPAKSPPG----------TSPPASRGA-PPLQAQPP  508

Qy   211  ITTTAKPINPRPSLPPNSDDYSKEPSLTVNKEHTTVETKQTSTDGKEKTTSAKETQ  270
Db   509  AASSPPATPVKSSPPPAAV-----------------------------------  527

Qy   271  SIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPTPKEPASTTTKEPTPTTI  330
Db   528  -VLPPPAKITPSSPPAV-ASPPPEEAPVSSPQPQVKSSPPPAPVASPPPPMKSP--PPPARV  583

Qy   331  KSAPTTPKEPAPTTPKEPAPTTPKSAPTTP-KEPAP--------TTTKEPAPTTPKEPAPTTTKEPAPT  382
Db   584  ASPPPLMKSPPPAPVASPPQPQLKSPPPPVLWLSTPSVKSPPPVPVPVASPPPPVKSPPPL  643

Qy   383  TTKSAPTTPKE----PAP--TTP----KKPAPTTPKEPAPTTPKE-PTPTTPKEPAPTTK  431
Db   644  APVSSPPSPPVKLPPLPAPGKSTPPPEEEKPTPPPTPVKSSPPPEEKSLPPPTLTTSPPPQEK  703

Qy   432  EPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPSPTTTKEPSPTTPKEPAPTT  491
Db   704  PTPPSTPSKPPPSPVETLPPPSKSSPPPBEPVSSPPQAPKSSSP-----PAPVS--SPPP  756

Qy   492  TTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEP  551
Db   757  LKSSPPPVPESSPPPTTPKSSPPLAPVSSPPQVEKTSPPPAPVSSPPPTTPKSSPPLAPVSS  816

Qy   552  APTTTKKPAPTTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT  611
Db   817  PPQVEKTSPPPAPVSSPPPTPKSSPPLAPVSSPPQVEKTSPPPAPVSSPPLEPKPSSPPS  876

Qy   612  PEEPTPTTPEEPAPTTPKAAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGT  671
Db   877  SVSSPPPTTVKSSPPPAPLSSPPMTPKSSSPPAHVSSPPEAFPKSSPPLAPISSPPSEPKSP  936

Qy   672  A--------PTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPT----TPKG  715
Db   937  SSPPMVEKTSPPPATVSSPPPTPKSSPPP-APVSSPPPVVKSSPPPAPVSSPPPTPKP  993

Qy   716  TAPTTP-------KEPAPTTPKEPAPTTPK--GTAPTTLKEPAPTTPKKPAPKELAPTT  765
Db   994  LPPPAPVSSPPPTPVSSPPPTPKPLPPPTPVSSPPPTPKPLPPP------APVS  1050

Qy   766  TKGPTSTTTSDKPAPTT--PKETAPTTPK----EPAPTTPKKPAPTTPETPPTTSEVSTP  819
Db  1051  SSPPPVVKSSPPPAPVSLPPPTPKRPSPPRTRVSSPRPVVKCCPPPTLVSSPPPAPKSLPPP  1110

Qy   820  TTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTIKTPAAIKPEMTTTAKDKTT  879
Db  1111  TPVSSP--------------PPEVKSSPPPPPVSSPPPAPKSSPPPTPVSSPPP------  1150
```

```
Qy   492  TTKEPAPTTPKSAPTTPKEP--SPTTPKEPAPTTPKEPAP---TTPPKKPAPTTPKEPAPT  546
Db   450  VAKPPVETPPIAPPPTAKPPISTPPISKPPVAPPAATPPITTTTPVKPPVATPPLAIPP  509

Qy   547  TPKEP--APTTTKPAPTAP--KEPAPTTPKETAPTT-----PKKLTP-TTPEKLAPT  594
Db   510  VAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVATPPVVTPPTTTPPTATPP  569

Qy   595  TPEKPAPTTPEELAPTT-PEEPPTTPTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPT  650
Db   570  VAKPPVATPPIATPPTAKPPISTPPISKSPVAPPAATPPITTPPAKPPVATPPIATPP  629

Qy   651  TPKEPAPTTPKETAP--TTPKGTAPTT--LKEPAPTTPKKPAPKELAPTTTKEPTSTTS  705
Db   630  IAKPPVATPPTATPPIATSPVAKPPVATPPTATPPIATPPVAKPPVAT--  687

Qy   706  DKPAPTTPKGTAP--TTPKEPAPTTPKEPAPTTPKKPAPTTPKKPAPKELAPT  764
Db   688  -PPTATPPIATPPIATPPVVTPPTATPPIATPPVTTI--PPTATPPVAMPIATPP  744

Qy   765  TTKGPTST--TSDKPAPTTPKETAPT-TPKEPAPTTPKKPAPTTP------EIPPPTTSE  815
Db   745  TAKPPIATPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVATPPIAKPP  804

Qy   816  VSTPTTTKEPTTIH--KSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAIKPEMI-T  872
Db   805  VATPPTIAPPTTATPPVAKPPVATPPIATPPTAKPPIILTPPISKPPVATPPAATPPITTPP  864

Qy   873  TAKDKTTERDLRTPTPETTTAAPKMTKETATTTEKTIATTTQVTSTTTQDTTPFKI  932
Db   865  PAKPPVVATPPIATPP--IAKPPVATPPTATPPIATSPVAKPPVAIPPICTKPQKPTKAPK  922

Qy   933  TTLKTTTLAPKVTTTKKT----ITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPK  988
Db   923  PPIATPPVAKPPVATPPTATPPIATSPIATPPVVT--PPTATSPVATPPIAKPPTTTP-  978

Qy   989  KPTSTKKPKIMPRVRKPKTPTPTRRKMTSTMPELNP------TSRIAE------AML  1032
Db   979  -PTAT-PPVAMPIATP---PTAKPPVATPPIANPPVEKPPVATPPIAKPPTVLPIPAKP  1033

Qy  1033  QTTTRPNQTPNSKL--VEVNPKSEDAGGAEGETPHMLLRHVFMPEVTPDMDYLPRVPNQ  1090
Db  1034  PVETSPTTATPPTATPPVAIPP-------VVKPPVAIPPITKPPVATPPVTNP  1078

Qy  1091  GIIINPMLS  1099
Db  1079  PTAMPPIVT  1087
```

```
Qy    880 ERDLRTTP------ETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTQDDTTPFKITT 934
              ::::|                      ::|:
Db   1151 --ELKSSPPAPVSSPPPSAPKSSSPPAPVSLPPPEVKSSSPPPAPISSPPPAPKSP---- 1203

Qy    935 LKTTTLAPKVTTTVKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTAPKPPSTK-- 994
              |||
Db   1204 -PPPA----------------------------------PMSSLPPPVKSPPPPAPVSS 1242

Qy    995 KPKTMPRVRKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENPKSE 1054
              ::|
Db   1243 SPPPAP--VKPPSLPPPAPVSSPPPAV----------------TSAP------ 1275

Qy   1055 DAGGAEGETPHHMLLRPHVHFMPEVTPDMDYLPRVPNQGIIINPMLSDE 1101
              ::|               |:|:  ::|
Db   1276 EDSTA----------------------------PPAEALPPSFENDILLPPIMANK 1303
```

RESULT 12

P70670

ID   P70670;   PRELIMINARY;   PRT;   2187 AA.
AC   P70670;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE   MUSCLE-SPECIFIC FORM GP220).
GN   NACA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96124550; PubMed=8698236;
RA   Yotov W.V., St-Arnaud R.;
RT   "Differential splicing-in of a proline-rich exon converts alphaNAC
RT   into a muscle-specific transcription factor.";
RL   Genes Dev. 10:1763-1772(1996).
DR   EMBL; U48364; AAB18734.1; -.
DR   EMBL; U48363; AAB18732.1; -.
DR   MGD; MGI:106095; Naca.
DR   InterPro; IPR002715; NAC.
DR   InterPro; IPR003037; TS-N.
DR   Pfam; PF01849; NAC; 1.
DR   Pfam; PF02094; TS-N; 1.
SQ   SEQUENCE 2187 AA; 220599 MW; 00346AA864DEBFD CRC64;

Query Match
Best Local Similarity   11.2%;
                        27.4%;
Matches 311; Conservative 128; Mismatches 449; Indels 246; Gaps 51;
Score 812;   Pred. No. 1.9e-44;   DB 11;   Length 2187;

```
Qy     68 NPISPPSSKKAPPSGASQTIKSTTKRSPKPNKKKTKVIESEEITEEHSVSENQESS 127
             ::|  ||||||
Db    746 SPT--PPSSKGAPVPSTGA----------------PPSPKGAPIVPIESSISSKQVPAEILPSQ 792

Qy    128 SSSSSSSSSTIWIKISSKNSAANRELQKKLKVKDNKNRTKKKPTPKPPVVDEAGSGLDN 187
             |:|
Db    793 KTPEVTASRLISAVQSPKVDPIMSD--------------VIPTSPKKTSATAVPK 833

Qy    188 GDFKVTTPDTSTTQOHNK-----VSTSPKITTAKPINPRPSLPPNSDTSKETSL--TVNKE 240
             |||
Db    834 --DTSATLSLKSVPAVTSLSPPKAPVAPSNEATIVPEIPSLKNALAAAPKE 885

Qy    241 T-------TVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKA 294
             |
Db    886 TLATSIPKVTSPSPQKIPKVSVSLKGAPAMTSKKAT---EIAASKDVSPSQ--FPKEVPLL 940

Qy    295 ETTTKGPALTTPKEPTPTPKEPASTTPKEPPTTIKSAPTTPKE-PAPTTTKSAPTTPK 353
             :||       :||:
Db    941 QHV---PPTSPPKSPVSDTLSGALTSPPPKGPPATLAETPTYPKKSSPKPAASKKTPATPS 997

Qy    354 ------EPAPTTKEPAPTTKEPAPTTTKEPAPTTT-TKSAPT-------------- 389
```

```
Db    998 PEGVTAVPLEIPPCSKRAPKTAAPKESSSATSSSKRAPKTAVSKEIPSKGVTAVPLEISLP 1057

Qy    390 ------TPKEPAPTTPKK-PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 440
             ||| ||    |||
Db   1058 LRETSKSATPGEKSASSPKRSPKTAGPKE---TPPGGVTAVPPEISLPPKETPQNATPNE 1114

Qy    441 PAPTAPKKPAPTT--PKEPAP-----TTPKEPAPTTTTKPEPSPTTPK-----EPAPTTKSA 489
             :||    ||     :||
Db   1115 SLAASSQKRSPKTSVPKETPGGGVTAMPLEIPSAPQOKAPKTAVPKQIPTPEDAVILAGS 1174

Qy    490 PTTTKEPAPTTT-KSAPTTPK------EPSPTTTKEPAPTTTPKEPAPTTPKEPAPTT- 539
             |||
Db   1175 PLSPKRASKTAAPKEAPATPSVGVIAVSGEISPSPKKTSRKTAAPKENSATLPPKRSPKTA 1234

Qy    540 -PKE------------PAPTTPKEPAP-----------TTTKKPAPTAPKEPAPTTPKETAPTT 579
             |||
Db   1235 APKETPATSSEGVTAVPSEISPSSPPTPASKGVPVTLTPKGAPNALAE-SPASPKKVPKTA 1293

Qy    580 PKKLPTTPKEKLAPTTPEK-PAPTTPEELAPTTPEEPTTTT----PEEPAPTTPKA 630
             :||     :|||
Db   1294 APEEITSTTP-----SPOKIPKVAGPKEASATPPSKKTPKTAVPKETSAPSEGVTAVPLE 1347

Qy    631 AAPNTPKEPAPTTPKE-PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTP--- 684
             :||
Db   1348 IPPSPRKAPKTAAPKETPAPS--PEGATTAPVQIPPSPRKGSKKAGSKE-TPTTSPEGV 1404

Qy    685 -------------KKPAPKELAPTTKEPTS--------TTSDKPAPTTPKGTAPTTPKE 723
             |||
Db   1405 TAAPLEIPISSKKTKSMASPKEILVTPSSKKLSQTVGPKEETSLEGATAVPLEIPPSHKKA 1464

Qy    724 PAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKEPAPKELAPTTTKGPTSTTSDKPAPTTP 782
             :||     |||
Db   1465 PKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-APKTAAPPSER-VTTVPPEKPA-TPQ 1519

Qy    783 KETAPTTPKEPAPTTPKKPAPTTPETP----PPTTSEVSTPTTTK-----EPTTIHKS 831
             |:|
Db   1520 KASGTTASKVPVPAETQEVAVSSRETPVTPAVPPVKNPSSHKKTSKITELKEAPATLPPS 1579

Qy    832 PDESTPELSAEPITPKALENSPKE-PGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETT 890
             |:|
Db   1580 PTKSPKIPSSKKAPRT--SAPKEPPASPSIK-PVTT--SLAQTAPPSLQKAPSTTIPKEN 1634

Qy    891 TAAPKMTKETATTTEKTTESKITATTTQVTSTTQDDTTPFKIITLKTTTLAPKVTTTVKKT 950
             |||      :|:
Db   1635 LAAPAV----LPVSSKSPAAPARASASLSPATAAPQTAPKEATTIPSCKKAAAATETPIET 1690

Qy    951 ITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKRKTMPRVRKPKTTPT 1010
             :|||
Db   1691 STAPSLEGAPKEISE----TSVSKVLMSSP-----PKKASSSKRASTLP------ 1730

Qy   1011 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAGGAEGSETP 1064
             :||
Db   1731 ------ATTLPSLKEASVLS------------PTATSSGKDSHISPVS-DACSTGTTTP 1770
```

RESULT 13

Q9VEL9

ID   Q9VEL9;   PRELIMINARY;   PRT;   2112 AA.
AC   Q9VEL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG4090 PROTEIN.
GN   CG4090.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;

```
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherly J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moshrefi A.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
SQ   SEQUENCE 2112 AA;  B91018E551A5D36 CRC64;

DR   EMBL; AE003716; AAF55402.1; -.
DR   FlyBase; FBgn0038492; CG4090.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF01607; Chitin_bind_2; 10.
DR   SMART; SM00494; ChtBD2; 11.
DR   PROSITE; PS00022; EGF_1; 1.

Query Match         10.8%;  Score 786.5;  DB 5;  Length 2112;
Best Local Similarity   26.9%;  Pred. No. 8.2e-43;
Matches 313;  Conservative 108;  Mismatches 380;  Indels 361;  Gaps 49;

Qy    21 QVSSQELSCKGRCFESFERGRECDCD-----------AQCKKYDKCCPDYESFCAEVHNPTSPP 73
          |:|                                    |
Db   953 QSSSNQTSSSTTQKPFKPAEKCESEETFLADNENCSKFYRCVDNGKGGFTKV-SFTCPP 1011

Qy    74 SSKKAPPPSGASQTIKSTTKRSRSRPKPPNKKTKKVI----ESEITEEHSVSENQESSSSS 129
          :|  |                                      |::| | :
Db  1012 NTIWDPEANSCNHPDQIQTK------PLKCKKVVSQGGSSSNSTSNSSSSSNNNSGSSSN 1064

Qy   130 SSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTPKKPPVVDEAGSGLNGD 189
          ||:||                                                     |
Db  1065 SGSSSSS------SSSNS--------------------------------GSSSNIGS 1084

Qy   190 FKVTTPDTSTTQHNKVSTSPKITTAKPINPRSLPPNSDTSKETSLIVNKETTVETKETT 249
          :                                                        |
Db  1085 SSNSGCASSSGGSSNQGSSS------------------------------------SS 1122

Qy   250 TTNKQTSTDGKEKTTSAKETQSIEKTSAK-DLAPTSKVLAKPTPKAETTTK-------- 299
          ||:|       :|| || :|  | : |:                    |
Db  1123 TSSSSSSSNNNQGSSSSSTSSSKPNPSETCKVNGQFIGDRSDCAKFYRCVDNDR 1182

Qy   300 ----------------GP--------------------ALITPKEPTPTTPKEPASTTPKEPTPT 328
                          ||                     |  ||  ||  ||   ||   ||
Db  1183 GGFNMVFPSCGPGTVWDAQMQACNHAWAVKECGGIAPPTTSTPPTTSR-PTTASTSRPSDQ 1241

Qy   329 TIKSAPTTPKEPAPTTVKSAPTTPKEPAPTTPKEPAPTTPKEPAPTT----------- 375
          | :|| ||  ||  || :|| ||  ||  ||  ||  ||  ||
Db  1242 TSTSRPTGP---PTTARPVTARPTTSSPTTASSSQTTSPVTQAPNTDGKCRSEGFMADP 1297

Qy   376 ------------------TKEP-----APT----------------TTKSAPTTPKEPAPTT 398
                            ||:|       ||                  | :|| ||  ||
Db  1298 NNCSKFYRCVRNNKGGFTSIPFQCGAGTVWDQDLQTCNHNFNNCSTGTGESTTPKEPA 1355

Qy   399 PKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTAPKKPAPTTPKEPA 458
            || ||  ||  ||  ||  ||  ||  ||  ||  ||         ||  ||  ||
Db  1356 PATNGTTATSTSSTTTD---PPTTTDLPPTSTTGLP-PTTTELPPT-----TTTDLPP 1405

Qy   459 PTTPKEPAPTTPKEPSPTTPKEPAPTTTTKSAPTTPKEPAPTTPKEPAPTTPK--EPSPTT 515
            || ||  ||  ||  ||  ||  ||   :|| ||  ||  ||  ||  ||    ||||
Db  1406 TTTRLPPTTTTSLPPTTTGLPPTTTTGAQPTTTTLSSETETSTVTISPESTTQPPSTT 1465

Qy   516 TKEPAP------------------------------------------------TTPK---E 526
          ||| |                                                 | |
Db  1466 TMKPLPAGTECTCGECYMAPDEDCRKYYRCINAGASYRKYNFTCPKGTGWNEEVQTCDYVE 1525

Qy   527 PAP-----TTPKPAPTTP------KEPAPTTPKE-PAPTTPTKKPAPTAPKEPAPTTPKETAPT 578
              |              |       ||  ||  ||  ||   ||  ||  ||  ||
Db  1526 NIPRCSKLPAEPITTTPSEESKDPGSTTPQSTDEPTTVTKPI-TKPTE-EPSTEKPQKPT 1583

Qy   579 TPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTPTEEPAPTTPKAAAAPNTPKE 638
          | ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  1584 TQYPEKPTTTEE--PEKPQKPT-TTEYPQKPTTTEEPTTEEYPQKPTTTEE 1639

Qy   639 PAPTTPKEPAPTT--PKEPAPTTPKETA------PTTPK---GTAPTTLKEPAPTTPKKPAPK 690
            || ||  ||      ||  ||  ||       | |        ||  ||  ||
Db  1640 PTPEKPQKPTTTEYPQKPTTEEPTTTSIPGYNPTTTSVPGYNPTTTPIPVETTTSTPGYK 1699

Qy   691 ELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKEPA 749
             ||  ||  ||  ||  ||  ||  ||  ||  ||  ||   ||  ||  ||
Db  1700 ---PTTTGEPTITTTT-LPSTTTDAIQEPTTSKKPEPTTTTESPESSTPEGSVTTLQPEPQ 1755

Qy   750 PTTTPKK-----PAPKELA--------PTTTKGPTSTTISDKPAPTTPKETAPTTPKEPAPTTP 798
           ||  ||      || :                ||  ||  ||  ||  ||  ||
Db  1756 PNYNCSSEGFPPDPEDCSRYYRCVDAAKNGKYQVYAFKCGKGTVWDTSTETCNYADQVSG 1815

Qy   799 KKPAPTTPETPPPTTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVP 858
           ||  ||  |||     ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  1816 N------------CSSGQPTTPCTTTEPGTTESTTSSGKP-----ETTSKAPEN 1852

Qy   859 TTKTPAATKPEMTTTAKKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKIIATTTTQ 918
           |||    |||   ||     | |      ||   ||  ||:|   |  ||   |
Db  1853 -TTTWA--PETTTT--------SSPETTTT--VASETTTTTSGT-----TTTA 1887

Qy   919 VTSTTTQDTTTPFKITTLKTTTLLAPKVTTTKTTITTEIMNKPEETAKPKDRATNSKATTP 978
          :|| ||     ||    ||   |   :|  ||  |  |  ||        |        ||
Db  1888 TPETTTKPPKP-----ETTTIAGEEPTSTKSPTTTE---------------------SP 1920

Qy   979 KPQKPTKAPKKPTTSTKPKTMP 1000
           |  ||  ||  ||  ||  |
Db  1921 APSTNTSAP------CPETGP 1935
```

RESULT 14
```
ID   Q41805      PRELIMINARY;    PRT;   1188 AA.
AC   Q41805;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
```

```
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
RA   EMBL; Z34465; CAA84230.1; -.
DR   Mendel; 14346; Zeama; 2368; 14346.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   SMART; SM00370; LRR; 4.
KW   Signal.
FT   SIGNAL        1     27       POTENTIAL.
SQ   SEQUENCE   1188 AA;  120980 MW;  2C77C7F8D7130149 CRC64;

Query Match             10.8%;  Score 784;  DB 10;  Length 1188;
Best Local Similarity   29.9%;  Pred. No. 6.6e-43;
Matches  230;  Conservative  59;  Mismatches 359;  Indels 122;  Gaps  33;

Qy    171 PTPKPPVVDEACSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSL----PPN 226
              |||| |         ||                     |  || |       |||
Db    460 PTPHSPPAD------DYVPPTPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQ 510

Qy    227 SDTSKETSLIVNKETTVETKETTTTINKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV 286
                                   |  |  |                |||
Db    511 APVG-------------------------SPPPPVKTTSPPAPIG----SPSPPPVSVV 541

Qy    287 -----LAKPTPKAETTTKGPALTTPKEPTPTTTPKEPTPTTTIKSAPTTPKEPA 341
                 : |      |||||           |||             |||
Db    542 SPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVVKSP--PPPTLVASPPPPVKSPP 599

Qy    342 PTTTKSAPTTP---KEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTIKSAPTTP-KEPAPTT 398
              :|| ||     |||             |||               |||       |||||
Db    600 PPAPVASPPPPVKSPPPPTPVASPPPPAPVASPPPPAPVASPPPPEKSPPPPP 659

Qy    399 PKKPAPTTPKEPAPTT----------PKE--PTPTTPKEPAPTTTKEPAPTTPKEPAPTTPKKP 449
              |||        |||            |||       |||             |||||  ||
Db    660 PAKSPPPPEEYPTPPTPVSKSSPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP-PSSPEKP 718

Qy    450 APTTPKEPAPTTPKEPAPTTTKEPSPTTTKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPK 509
              |||||             |||             |||             |||
Db    719 SP--PREPVSSPPQTP----KSSPPPAPVSSPPPPTPVSSPPALAPVSSPPPSVKSSP---- 768

Qy    510 EPSPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTTKKPAPTAPKERAP 569
              ||             |||             |||               ||
Db    769 PPAPLSSPPPAPQVKSSPPPVQVSSP--PPAPKSSPPPLAPVSSPPQVEKTPSPPAPLSSPP 827

Qy    570 TTPKETAP----TTP----KKLTPTTPEKLAPTTPEKPAP------TTPEELAPTTPEEPTT 616
              |||       |||       :|                         :||
Db    828 LAPKSSPPHVVSSPPPPVKSSPPPVKSSPPPPALTPKPASPPAHVSSPPEVVKPSTP--PA 885

Qy    617 PTTPEEPAPTTPKAAANPKE-PAPTTPKEPAPTTTPKEPAPTTPKETAPTTPKGTAPTT 675
              |||: | |  | ||::  |||  | || ||             ||| |      |||| |
Db    886 PTTVISP-PSEPKSSPPPPTPVSLPPPIVKSSPPPAMVSSP-PMTPKSSPPPPVVVSSPPT 943

Qy    676 LKE-------PA------PTTPKK---PAPKELAPTTTKE--PTSTTSDKPAPTTPKGTAPTTP 721
              |||        |          |||              |||
Db    944 VKSSPPPPAPVSSPPATPKSSPPPAPVNLPPPPEVKSSPPPTTPVSSSPPPA--PKSSPPPAP 1000

Qy    722 -KEPAPTTPKGTAPTTPKPAPKELAPTTTTKGPTSTTSDKPAPTT 780
                |||             |||             |||
Db   1001 MSSPPPPEVKSPPPPAPVSSPPPPAPVSSSPP---PPVKSPPPPAPVSSSPPPPV 1057

Qy    781 -TPKETAPTTPKEPAPTTTPKTPEPTTPEPTTSEVSTPTTTKEPTTIHKSPDESTPEL 839
              ||:|||             |||             ||             |||
Db   1058 KSPPPPAPISSSPPPAVKSSPPPAVSS--SPPPEVKSPPPPAPVSSSPPPPIKSPPPAPVS 1115

Qy    840 SAEPTPKRALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTTERDLRTTPET 889
              |                  |||           |     |
Db   1116 SPPPAP------VKPPSLP-PPAPVSSSPPPVVTPAPPKKEEQSLPPPAES 1158
```

```
RESULT   15
ID   Q9XDH2                PRELIMINARY;                  PRT;      763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH MUCIN HOMOLOG.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang Y., Moreno C., Singh M.;
RT   "Cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003882; Pistil_extensin.
DR   PRINTS; PRO1222; ATROPHIN.
DR   PRINTS; PRO1217; PRICHEXTENSN.
DR   PRINTS; PRO1218; PSTLEXTENSIN.
SQ   SEQUENCE   763 AA;  75034 MW;  39168EC45A5916F8 CRC64;

Query Match             9.8%;  Score 715;  DB 2;  Length 763;
Best Local Similarity   28.2%;  Pred. No. 1.3e-38;
Matches  248;  Conservative  39;  Mismatches 361;  Indels 232;  Gaps  39;

Qy    290 PTPKAETTTKGPALTTPKEPTPTTTPKEPASTTTPKEPTPTTTIKSAPTTPKEPAPTTTKSAP 349
              |||           |||             |||             |||
Db      3 PVP-------APRALAPLPPAPPAPAEPRKSKPPPPAP-------PAPPCWMLVSAAP 46

Qy    350 TTTPKEPAPTTTPKEPAPTTTKEPAPTTTTPKEPAPTTTPKEPAPTTTKKPAPTTPKE 409
              |||             |||             |||
Db     47 PCP--PAP-----PAPPKPKSKAPPPPVPPAPPARELAPPLP--PAP-------PEAPRE 90

Qy    410 PAPTTPKEPTPTT----PKEPAPTTKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK 463
              |||             |||             |||             |||
Db     91 SRPALPPCPPPPVVIPDPEPEAAPPVPPAPNSPPFPFPPPPAPKFVPAPPVP--PVPNSPP 148

Qy    464 EP--APTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTPKEPSPTTTKEPAP 521
              ||               |||             |||             |||
Db    149 FPPPPPAALNPPAP-----PAPPLANSPPLPPAPPTPAGT--PPAAPWPPVPAAPKSKPA 201

Qy    522 TTTPKEPAPTTPKEPAPTTTPKEPAPTTPKEPAP---TTTKKPAPTAPKEPAPTTPKETAP 577
              |||             |||             |||             |||
Db    202 SPPRPPAP-----PMPATPMEFPPLPPVPPPDPISKETPAPPAPPIPAPVPIPVPPLP 256

Qy    578 TTTPKKLTPTTPEKLAP------------TTPEKPAPTTPEELAPTTPEEPTTPTTPEEPAPTTT 627
              |||             |||             |||
Db    257 PVPNKIPPAPP--APPVAVAAVLVAPCPLPLPLPNNHPPAPPAAAPAPVPGVPLAPLPNSHP 313

Qy    628 ---PKAAAPNTPKEPAP---------- 641
                |||
Db    314 PAPPSAPVPGVPLAPLPLSGRPVSVWKGSFTTLSTFCCRVCSGEVLAGALNPSRPSRSPL 373

Qy    642 --TTPKEPAPTTPKEPAPTTTPKETA----PTTPKGTA---PTTLKEPAPTTP-KPAPKE 691
              ||               |||             |||             |||
Db    374 TTTTPALPAPIPLPLPLPPLPLFINTAVPPIPLPPVTALAPPLPPLAPLPISPGVPAPPAP-- 431

Qy    692 LAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPA-----PTTPKEPAPTTPKGTAPTTL 745
              |||             |||             |||
Db    432 --PIPGKPWTTPPLAPAPPEPEK-TVPVLPPGSCPPSEKPNPPAPPEEPPERKSSPPALPP 488

Qy    746 KEPAPTTP---KKPAPKELAPTTTKGP-TSTTSDKPAPTTPKET---APTTPKEPAPTTP 798
              |||             |||             |||             |||
```

```
Db   489 APPAPSMPSAVRPPSPPIPPAPPAAPRASMPALPPAPPSPPATRLCPPLPPSPPAPNSP 548

Qy   799 KKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSP----DESTPELSAEPTPKA--LENSP 852
         |||  : ||| :||||||| ||  :|||
Db   549 --PAPPAPPTPPKLLS--ANPPCPVPPAPNRPPAPPAPPELPAPPDPPTPPVANSP 604

Qy   853 KEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTTESKI 912
         |   :  :        | ||||
Db   605 --PAPPAPPSALPFVNPPA--------------PPTPAAPK-----------SRPAL 637

Qy   913 TATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATN 972
         |
Db   638 PAAPPAPPAPPVRATTP----------------------PPAPPAPPAPN 665

Qy   973 SKATTPKPQKP----TKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTM-----PELNP 1023
         | : | |  :  | :||      |:                           |: |
Db   666 SMALPPAPPDPPIPLLATPPAPPAPPLPMSPPAPPLPPAAPDPPAPPLTINQPPSPPLAP 725

Qy  1024 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET 1063
         |  |          | : |:|   | :|:|
Db   726 VPGAPLAPLPINGRPVFARKNSLI--------GSSSGDT 756
```

```
                GenCore version 4.5
                Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:         April 26, 2002, 16:29:14 ; Search time 210.7 Seconds
                                              (without alignments)
                                              881.660 Million cell updates/sec

Title:          AA3
Perfect score:  6814
Sequence:       1 MAWKTLPIYLLLLSFVIQ.........ARAITTRSGQTLSKVWYNCP 1270

Scoring table:
Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters:       473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SPTREMBL_17:*
                1: sp_archea:*
                2: sp_bacteria:*
                3: sp_fungi:*
                4: sp_human:*
                5: sp_invertebrate:*
                6: sp_mammal:*
                7: sp_mhc:*
                8: sp_organelle:*
                9: sp_phage:*
               10: sp_plant:*
               11: sp_rodent:*
               12: sp_virus:*
               13: sp_vertebrate:*
               14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                       SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6727	98.7	1404	4	Q92954	Q92954 homo sapien
2	6712	98.5	1404	4	Q9BX49	Q9bx49 homo sapien
3	3369	49.4	1054	11	Q9JM99	Q9jm99 mus musculu
4	1713.5	25.1	401	6	O77765	O77765 bos taurus
5	971.5	14.3	1079	5	Q9N4S7	Q9n4s7 caenorhabdi
6	936.5	13.7	1049	5	Q9I7S1	Q9i7s1 drosophila
7	865.5	12.7	1795	5	O76894	O76894 drosophila
8	855	12.5	1489	10	O96449	O96449 phytophthor
9	811.5	11.9	1315	10	Q9SPM0	Q9spm0 zea mays (m
10	802	11.8	1274	5	Q20007	Q20007 caenorhabdi
11	800.5	11.7	1480	10	Q9IE8	Q9iie8 arabidopsis
12	784	11.5	1188	10	Q41805	Q41805 zea mays (m
13	776.5	11.4	2187	11	P70670	P70670 mus musculu
14	756	11.1	2112	5	Q9VEL9	Q9vel9 drosophila
15	715	10.5	763	2	Q9XDH2	Q9xdh2 mycobacteri
16	707.5	10.4	555	10	Q9FPQ6	Q9fpq6 chlamydomon
17	671	9.8	1151	13	O57580	O57580 gallus gall
18	670	9.8	1229	5	Q94185	Q94185 caenorhabdi
19	668	9.8	2284	5	Q9VPG1	Q9vpg1 drosophila
20	667.5	9.8	4880	11	Q9LLT1	Q9llt1 rattus norv
21	667.5	9.8	5085	11	Q9JKS6	Q9jks6 rattus norv
22	667	9.8	6677	5	Q9N435	Q9n435 caenorhabdi
23	662	9.7	2089	4	Q14676	Q14676 homo sapien
24	653.5	9.6	1514	5	Q9GUW7	Q9guw7 leishmania
25	651.5	9.5	3570	4	Q99552	Q99552 homo sapien
26	647	9.5	4833	11	Q9QX6	Q9qyx6 mus musculu
27	647	9.5	5038	11	Q9QYX7	Q9qyx7 mus musculu
28	642.5	9.4	1612	5	Q9VYQ2	Q9vyq2 drosophila
29	641.5	9.4	3507	5	Q23587	Q23587 caenorhabdi
30	635.5	9.3	7962	4	Q10465	Q10465 homo sapien
31	635	9.3	2768	5	Q9VC00	Q9vc00 drosophila
32	633	9.3	489	10	Q41707	Q41707 vigna ungui
33	632	9.3	761	10	Q92QIO	Q92qio arabidopsis
34	632	9.3	6632	5	O17362	O17362 caenorhabdi
35	630.5	9.2	990	13	Q91803	Q91803 xenopus lae
36	626.5	9.2	971	5	Q9XVS4	Q9xvs4 caenorhabdi
37	625	9.2	6642	5	O01761	O01761 caenorhabdi
38	622.5	9.1	839	2	Q9RX57	Q9rx57 deinococcus
39	615	9.0	2344	5	Q9N3Y8	Q9n3y8 caenorhabdi
40	608	8.9	801	5	Q23635	Q23635 caenorhabdi
41	607.5	8.9	924	12	Q99307	Q99307 epstein-bar
42	605	8.9	379	5	Q27929	Q27929 drosophila
43	600.5	8.8	1893	5	Q9NKC9	Q9nkc9 drosophila
44	598.5	8.8	409	10	Q9SBM1	Q9sbm1 volvox cart
45	577	8.5	956	10	Q9LJ64	Q9lj64 arabidopsis

```
                                      ALIGNMENTS

RESULT     1

ID     Q92954                  PRELIMINARY;       PRT;    1404 AA.
AC     Q92954;
DT     01-FEB-1997 (TrEMBLrel. 02, Created)
DT     01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT     01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE     MEGAKARYOCYTE STIMULATING FACTOR.
OS     Homo sapiens (Human).
OC     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX     NCBI_TaxID=9606;
RN     [1]
RP     SEQUENCE FROM N.A.
RA     Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA     Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.;
RL     SEQUENCE FROM N.A.
RN     [2]
RP     SEQUENCE FROM N.A.
RA     Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA     Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA     FitzGerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA     Jacobs K., Turner K.;
RL     (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL     Mosher D.F. (eds.);
RL     Biology of vitronectins and their receptors., pp.45-52,
RL     Elsevier Science Publishers B.V. (1993).
RN     [3]
RP     SEQUENCE FROM N.A.
RA     Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA     Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,
RA     Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA     Bhatia S., Kriz R., Hewick R., Clark S.C.;
RA     Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA     Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL     Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RL     Blood 78:279-279(1991).
RL     Blood 78:279-279(1991).
DR     EMBL; U70136; AAB09089.1; -.
DR     InterPro; IPR000585; Hemopexin.
DR     InterPro; IPR001212; Somatomedin_B.
DR     InterPro; IPR002400; GF_cysknot.
DR     Pfam; PF00045; hemopexin; 2.
DR     Pfam; PF01033; Somatomedin_B; 2.
DR     PRINTS; PR00438; GFCYSKNOT.
```

```
DR   PRINTS; PR000222; SOMATOMEDINB.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
SQ   SEQUENCE   1404 AA;   151090 MW;   AABD7AD19B35F4F6 CRC64;

Query Match
Best Local Similarity   90.5%;          Score 6727;   DB 4;   Length 1404;
Matches 1270;  Conservative   0;  Pred. No. 0;   Mismatches   0;   Indels  134;   Gaps   2;

Qy      1 MAWKLPIYLLLLLSVFVIQQVSSQ-----------------------------       25
          ||||| ||||| |||||||||||||
Db      1 MAWKTLPIYLILLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF   60

Qy     26 -----ELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAE-------        65
               |||||||||||||||||||||||||||||||||||||||
Db     61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP  120

Qy     66 -------------------------------------------                 65
Db    121 PPSGASQTIKSTIKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSTIW  180

Qy     66 --VRDNKKNRTKKKPTPKPKPPVVDEAGSGLDNGDFKVTPDTST                106
          ||||||||||||||||||||||||||||||||||||||||||
Db    181 KIKSKSNSAANRELQKKLKVKDNKNRTKKKPTPKPPVVDEAGSGLDNGDFKVITPDTST   240

Qy    107 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG  166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDG  300

Qy    167 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPEPTTPKEPAS  226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPEPTTPKEPAS  360

Qy    227 TTPKEPTPTTIKSAPTTPKEPAPTTPKKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEP  286
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTPKEPTPTTIKSAPTTPKEPAPTTPKKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEP  420

Qy    287 APTTTKSAPTTPKEPAPTTPKKPAPTTPKGTAPTTPKGTAPTTPKEPAPTTPKEPAPTTPK  346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 APTTTKSAPTTPKEPAPTTPKKPAPTTPKGTAPTTPKGTAPTTPKEPAPTTPKEPAPTTPK  480

Qy    347 EPAPTTAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKSAPTTTKEPAPT          406
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 EPAPTTAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKSAPTTTKEPAPT          540

Qy    407 TTKSAPTTPKEPSPTTPKEPAPTTPKKPAPTTPKKPAPTTPKEPAPTTTKKP          466
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTKSAPTTPKEPSPTTTKEPAPTTPKKPAPTTPKKPAPTTPKEPAPTTTKKP          600

Qy    467 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPEELAPTTPEEPPTT           526
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPEELAPTTPEEPPTT           660

Qy    527 PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEP          586
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP          720

Qy    587 APTTPKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKG           646
          |||||||||||||||||||||||||||||||||||||||
Db    721 APTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKGTAPTTPKG           780

Qy    647 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKEPAPTTPK           706
          ||||||||||||||||||||||||||||||||||||||||
Db    781 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKEPAPTTPK           840

Qy    707 KPAPTTPETPTPPPTSEVSTPTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT   766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 KPAPTTPEPPPPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPT    900

Qy    767 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV  826
```

```
Query Match
Best Local Similarity   90.2%;          Score 6712;   DB 4;   Length 1404;
Matches 1267;  Conservative   0;  Pred. No. 0;   Mismatches   3;   Indels  134;   Gaps   2;

Qy      1 MAWKLPIYLLLLLSVFVIQQVSSQ-----------------------------       25
          ||||| ||||| |||||||||||||
Db      1 MAWKTLPIYLILLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF   60

Qy     26 -----ELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAE-------        65
               |||||||||||||||||||||||||||||||||||||||
Db     61 KRVCTAELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSSKKAP  120

Qy     66 -------------------------------------------                 65
Db    121 PPSGASQTIKSTIKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSTIR  180

Qy    887 PQKPTKAPKKPTSTKKRKTMPRVRKPKITTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN  946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQV   960

Qy   1021 PQKPTKAPKKPTSTKKPKITMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1006 QTPNSKLVEVNPKSEDAGSAEGETPHMLLRPHVPMPEVTPDMDYLPRVPNQGIIINPMLS  1006

Qy   1067 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGGLTGQIVAALSTAKYKNWPESVY 1126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 DETNICNGRKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDIVFT  1066

Qy   1187 SPARLAYQDRGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQIYNIDV   1246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYYNIDV   1380

Qy   1247 PSRTARAITTRSGQTLSKVWVNCP                                     1270
          |||||||||||||||||||||||
Db   1381 PSRTARAITTRSGQTLSKVWVNCP                                     1404
```

```
Qy     66 ------------------VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST 106
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 KIKSSKNSAANRELQKLKVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST 240

Qy    107 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLITVNKETTVETKETTITNKQTSTDG 166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLITVNKETTVETKETTITNKQTSTDG 300

Qy    167 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPTTPKEPAS 226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 KEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPTTPKEPAS 360

Qy    227 TTPKEPTTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEP 286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTPKEPTTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEP 420

Qy    287 APTTTKSAPTTPKEPAPTTPKKPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTPKEPAPTTK 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 APTTTKSAPTTPKEPAPTTPKKPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTPKEPAPTTK 480

Qy    347 EPAPTAPKKPAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKGPAPTTTKSAPTTTKEPAPT 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 EPAPTAPKKPAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKGPAPTTTKSAPTTTKEPAPT 540

Qy    407 TTKSAPTTPKEPSPTTTKEPAPTTTKEPAPTTTKKPAPTTPKKPAPTTTPKEPAPTTTKKP 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTKSAPTTPKEPSPTTTKEPAPTTTKEPAPTTTKKPAPTTPKKPAPTTTPKEPAPTTTKKP 600

Qy    467 APTAPKEPAPTTPKETAPTTTPKKLTPTTPKEKLTPTTPEELAPTTTPEEPTPTT 526
          ||||||||||||||||||||||||||||||||||||||||||||
Db    601 APTTTPKEPAPTTPKETAPTTTPKKLTPTTPEKLTPTTPEEKLAPTTTPEEPTPTT 660

Qy    527 PEEPAPTTPKAAAANTPKEPAPTTTKEPAPTTTPKEPAPTTTKGTAPTTLKEP 586
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    661 PEEPAPTTPKAAAANTPKEPAPTTTKEPAPTTTPKEPAPTTTKGTAPTTLKEP 720

Qy    587 APTTLKEPAPTTPKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTTPKEPAPTTPKG 646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 APTTTPKKPAPKELAPTTTKEPTSTTCDKPAPTTPKGTAPTTTPKEPAPTTPKG 780

Qy    647 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTTSDKPAPTTPKETAPTTPKEPAPTTPK 706
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TAPTTLKEPAPTTPKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTTPKEPAPTTPK 840

Qy    707 KPAPTTPETPPPTTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTTPKALENSPKEPGVPT 766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 KPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTTPKALENSPKEPGVPT 900

Qy    767 TKTPAATKPEMTTTAAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTTESKITATTTQV 826
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TKTPAATKPEMTTTAAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTTESKITATTTQV 960

Qy    827 TSTTTQDTTPFKITTLKTTTLAPKVTTTEIMNKPEETAKPRDRATNSKAATTPK 886
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TSTTTQDTTPFKITTLKTTTLAPKVTTTEIMNKPEETAKPRDRATNSKAATTPK 1020

Qy    887 POKPTKAPKPTSTKKPKTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 POKPTKAPKPTSTKKPKTMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080

Qy    947 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIINPMLS 1006
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 QTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIINPMLS 1140

Qy   1007 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPFPSPARRITEVWGIPSPIDTVFT 1066
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 DETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPFPSPARRITEVWGIPSPIDTVFT 1200

Qy   1067 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESSVY 1126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 RCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAALSTAKYKNWPESSVY 1260

Qy   1127 FFKRGGSIQQYIKQEPVQKCPCGRRPALNYPVYGEMTQVRRRFERAIGPSQTHTIRIQY 1186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 PSRTARAITTRSGQTLSKVWWNCP 1404
```

```
Db   1261 FFKRGGSIQQYIYKQEPVQKCPCGRRPALNYPVYGETTQVRRRFERAIGPSQTHTIRIQY 1320

Qy   1187 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGIDYYAFSKDQYYNIDV 1246
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 SPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGIDYYAFSKDQYYNIDV 1380

Qy   1247 PSRTARAITTRSGQTLSKVWWNCP 1270
          ||||||||||||||||||||||||
Db   1381 PSRTARAITTRSGQTLSKVWWNCP 1404
```

```
RESULT   3
Q9JM99

ID   Q9JM99          PRELIMINARY;    PRT;   1054 AA.
AC   Q9JM99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE   PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN   PRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ikegawa S., Nakamura Y.;
RT   "a novel mouse gene highly homologous to a human gene encoding
RT   megakaryocyte stimulating factor precursor and cartilage superficial
RT   zone protein.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB034730; BAA92310.1; -.
MGD; MGI:1891344; Prg4.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001212; Somatomedin_B.
Pfam: PF01033; Somatomedin_B; 2.
PRINTS; PR01217; PRICHEXTENSN.
PRINTS; PR00022; SOMATOMEDINB.
SMART; SM00120; HX; 2.
SMART; SM00201; SO; 2.
PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
PROSITE; PS005524; SOMATOMEDIN_B; 2.
SEQUENCE  1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match       49.4%;  Score 3369;   DB 11;   Length 1054;
Best Local Similarity  49.7%;  Pred. No. 2e-218;
Matches   701; Conservative   54; Mismatches   158;  Indels 498; Gaps 23;
```

```
Qy      1 MAWKTLPIYLLLLLSVEVIQQVSSQ------------------------------ 24
          | ||| |||||||||||||||||||||
Db      1 MGWKILPVCLSLLLPVVLIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60

Qy     26 -----ELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYESFCAE----------- 65
               |||| |||||  |:|||||||||||::::||:||:|:||:||
Db     61 KRVCSPELSCGRCFESFARGRECDCDSQCKQYGKCCADYDSFCEEVHNSTSPSSKTAPT 120

Qy     66 ------------------------------------------VKDNKKNRTKKK 65
                                                      |||||||||||
Db    121 PAGASDIKSTTKRSPKSPTTRTIKVVESEELTEHSDSENQESSSSSSSSSSSTIRKIKS 180

Qy     66 ------------------VKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVT---TPDTSTTQH 109
                            |||||||||||  ||: ||||:|:|||||||||||||    ||||||||
Db    181 SKNSANRELQKNPNVKDNKKNTPKKKPNPEPAVDEAGSGLDNGEFKLTPPPPDPPTTPH 240

Qy    110 NKVSTSPKITTAKPINPRPSLPPNSDTSKETSLITVNKETTVETKETTITNKQTSTDGKEK 169
          :|:|||||||||||||||||||||||||||||| | ||:||||:||||||||||| |:
Db    241 SKVATSPKTTAAKPVTPKSLAPNSETSKEASLASNKETTVETKETTATNKQSSA-SKKK 299

Qy    170 TTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPTTPKEPASTTP 229
```

```
Db  1024  QYYNIDVPTRTARAITTRSGQTLSKIWYNCP 1054

RESULT   4
         O77765
ID   O77765       PRELIMINARY;      PRT;   401 AA.
AC   O77765;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ARTICULAR CARTILAGE;
RA   Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA   Aydelotte M.B.;
RT   "Immunodetection and Partial cDNA Sequence of the Proteoglycan,
RT   Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT   Joints.";
RL   J. Orthop. Res. 0:0-0(1998).
DR   EMBL; AF056218; AAD13404.1; -.
DR   InterPro; IPR000585; Hemopexin.
DR   Pfam; PF00045; hemopexin; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   SMART; SM00120; HX; 2.
FT   NON_TER       1
FT   NON_TER     401
SQ   SEQUENCE   401 AA;  44952 MW;  86147CC9AFBB73D7 CRC64;

Query Match             25.1%;  Score 1713.5;  DB 6;  Length 401;
Best Local Similarity   80.2%;  Pred. No. 1.3e-107;
Matches  320;  Conservative  24;  Mismatches  52;  Indels  3;  Gaps  3;

Qy   873  PKDRADTNSKAITTPKPQKPTPKAPKKPTSTKKPKTMPRVRKPKTTPPTPRK-MTSTMPELNPT 931
          : :||||||||:|||:||||||||||:|||||:|||||:| |||||||||||| ||:||:|||
Db     5  PKGRATNSQVTTPKPQKPTKAPKKPTSTKKPRT-PRVRKPKITTPTPPKTTTSAMPEPTPT  63

Qy   932  SRIAEAMLQTTTPRPNQTPTNSKLVEVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYL 991
          :| ||||||||||||||:|||:::||:|||||||||||||||||||||||||||||:
Db    64  S-LPEAMLQTTTPRPTTPTPNSEIIDVNSENEDGDAAEGEKPHMIRPPVLTPIVIPGTEII 122

Qy   992  PRVPNQGIIINPMLSDETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPSPARRI 1051
          ||||||||||||:||:|||||||||:|||||||||||||||||||||||||||:|||||||
Db   123  VRGPSQGFGINPMFSDDTNLCNGRPVDGLTTLRNGTLVAFRGHYFWMLIPFTPPPPPPRRI 182

Qy  1052  TEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFFNDIKDAGYPKPIFKGFGGLTGQIVA 1111
          ||||||||||||||||||||||||:|||||||||||||||||||||||||||:||||||||
Db   183  TEVWGIPSPIDTVFTRCNCEGKTFFFKDSQYWRFFNDIKDAGYPKLISKGFGGLNGKIVA 242

Qy  1112  ALSTAKYKNWPESVYFFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFE 1171
          |||||||:|||::|||||||||||||||:||:|||:||||||:|||:||||||||||||||
Db   243  ALSIAQYKSRPESVYFFKRGGSVQQYTYKQEPTQKCTGRRPAINYSVYGETAQVRRRFE 302

Qy  1172  RAIGPSQTHTIRIQYSPARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGY 1231
          |||||||||||:|||||||||||||||||||||||||:||||||||||||||||||||||
Db   303  RAIGPSQVHTIRIHYTPVRVPYQDKGFLHNEVKVSTMWRGFPNVVTSAITLPNIRKPDGY 362

Qy  1232  DYYAFSKDQYYNIDVPSRTARAITTRSGQTLSKIWYNCP 1270
          |||||||||||||||||||||||||||||||||||||||
Db   363  DYYALSKDQYYNIDVPSRTARAITTRSGQTLSNTWYNCP 401

RESULT   5
         Q9N4S7
ID   Q9N4S7       PRELIMINARY;      PRT;   1079 AA.
AC   Q9N4S7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```

```
Qy   300  TTSVKETRSAEKTSDKDV----------------------------------------EPTSTTP 324
          ||| ||||:|||:|||| |:                                      || ||||
Db   230  KEPPTTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 289

Db   325  K------NSAPTTKKPV-TTTKESKFLP---------LPQEBEPPTTAKEPPPT 362
          |                |||:||:| :|:         |:|| |||||||| |||
Qy   290  TTKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPG 349

Qy   363  TKKPEPTTPKKEPEPTTPKKEPEPTTPKKEPEPTTPKEPEPTTPKEPEG 422
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   363  PTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKEPAPTTTK 409

Db   410  SAPTTPKEPSPTTPKEPAPTTPKEPAPTTPKKPAPTTPKKPAPTTPKKPAPT 461
          |||||||| |||||||||||||||||||:|||| ||| ||  ||
Qy   423  PTTTPKEPEPTTPKEPEPTTPKEPESTTRKEPEPTTPKEP 461

Qy   462  -EPTTPKEPEPTTLKEPEPTTPKEPEPTTPKEPEPTTPKEPEPPTTPKEPEPT----- 512
          : |||:||||||| ||||||||||||||||||||||||:|||||| || ||:::
Db   470  APKEPAPTTPKETAPTTPKKLTPTTPKKLAPTTPKEPAPTTPEELAPTTPEEPTPTTPEE 529

Db   530  PAPTTPKAAAPNTPKEPAAPTTPKEPAPTTPKGTAPTTLKEPAPT 589
          ||||||||  || |||  |||:|||:||| || |:|||||||||
Qy   513  --TPKEPEPTTLKEPEP---------------------------------TTKEPEPTTKEP 512

Qy   513  TPKKP-----------EPTTKEPVPTTPKEPEPTTP----- 570
          ||||||          |||||||||||||||||||||
Db   545  PAATKPEMTTPAKDKTTERDLRTTPETTTAAPKMTKETATTTEKITTESKITATTTQVTST 829

Db   590  TPKKPARPKELAPTTTKEPTSTTSDKRAPTTPKGTAPTTPKGTAPTTPKEPAPTTPKGTAP 649
          ||||||||||| ||||||||:||| ||||||| ||| ||| ||| ||||||||||| |||
Qy   545  -------EPTTPKEPVPTTPKEPEPTTP----- 570

Qy   615  PTT-------------------KEREPT 544
          |||                  |||:|||
Db   710  PTTPEPEPPTTSEVSPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKT 769

Qy   618  ---------------TSPKT--------- 622
          :::|                :::||
Db   770  PAATKPEMTTPAKDKTTERDLRTTPETTTAAPKMTKETATTTEKITTESKITATTTQVTST 829

Db   830  TTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETTAKPKDRATNSKAT----- 883
          ||| ||||:||||||||| |||:||||||||:||||:
Qy   623  --TPKP-OKPTKAPKKPTSTKPKPKTMPRVRKPKTTPPTPPRKMTSTMPELNPTSRIAEAMLQ 667

Qy   650  TTLKEPAPTTPKKPAPKELAPTTTTKGPTSTTSDKRAPTTPKETAPTTPKEPAPTTPKKPA 709
          | :|||||||| ||||||:|||||||||||||:|||||||||||||||||||||||||
Db   884  -TPKP-OKPTKAPKKPTSTKPKPKTMPRVRKPKTTPPTPPRKMTSTMPELNPTSRIAEAMLQ 940

Db   668  TKRAPKPKKPTKAPKKPTSTKKPKT--PKTRKPKTTPAPLKTTSATPELNTTP--LEVMLP 724
          |: ||:||||||||:|||||||||||  |: ||||||||||:||| |||||||   |:|||:
Qy   571  ---KEREPTTPKEPEP--------------------------TTKEPEPTTPKEPEP 614

Qy   725  TTTIPKQTPNPETAEVNPEDHEDADGEGEKP-LIPGPPVLFPTAIPGTDLLAGRLNRGIN 783
          ||||:||:||||| ||:||||||||:|||||  |||||| | |  ||:||| || | |||
Db  1001  INPMLSDETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSP 1060

Db   784  INPMSDETNLCNGKPVDGLTTLRNGTLVAFRGHYFWMLNPFRPPSPPRITEVWGIPSP 843
          |||:|||||:|||||||||||||||||||||||||||| ||:||||||||||||||||
Qy  1061  IDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAAISTAKYKN 1120

Qy   844  IDTVFTRCNCEGKTFFFKDSQYWRFTNDVVDPGTPKQIVKGFGGLTGKIVAALSIAKYKD 903
          ||||||||||||||||||||||||||||:|  | |||::||||||||:||||||||||||
Db  1061  IDTVFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVAAISTAKYKN 1120

Db   904  RPESVYFFKRGGNIQQYYKQEPMKKCIGRRPAINYSVYGEAAQVRRRFERAVGPFQTH 963
          ||||||||||| ||||:|||||:|||:|||||||||||||  |||||||| ||| |||
Qy  1121  WPESVYFFKRGGSIQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRFERAIGPSQTH 1180

Qy  1181  TIRIQYS-PARLAYQDKGVLHNEVKVSILWRGLPNVVTSAISLPNIRKPDGYDYYAFSKD 1239
          |||:|||  ||:|:|||| |||||||||::|||:||||||| ||||||||||||||| ||
Db   964  TFRIHYSVPMRVSVQDKGFLHNEVKVSTMWRGFPNVVTSAITLPNIRKPDGYDYYAFSKD 1023

Qy  1240  QYYNIDVPSRTARAITTRSGQTLSKVWYNCP 1270
          |||||||||||||||||||||||||:|||||
Db  1024  QYYNIDVPSRTARAITTRSGQTLSKIWYNCP 1054
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Y51B11A.1 PROTEIN.
OS   Y51B11A.1.
GN   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Cotton M.;
RT   "The sequence of C. elegans cosmid Y51B11A.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC0067797; AAF60743.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match    14.38%;  Score 971.5;  DB 5;  Length 1079;
Best Local Similarity   29.5%;
Matches  330;  Conservative 129;  Mismatches 481;  Indels 177;  Gaps 44;
```

```
Qy    65 EVRDNKNKNRTKKKPPTPKPPVVDEAGSGLDNGDFKV------TTPDTSTTQ-HNKVSTSP 116
            :||  : |:| ||
Db    31 EIADCRANQTPHTMLP----------STLTSVDMETPSTLVLSSTPTSSSTPIKETTTTAP 81

Qy   117 KITTAKPINPRPSLPPNSDTSKETSLTVNKETT-VETKETTTTNKQTSTDGKEKTTSAKE 175
Db    82 ETTSTEP-------PSSSTTPVQTTTTTAPETTSTEAPSSSSTTPVQTT------TTTAPE 128

Qy   176 TQSIEKTSAKDLAPTSKVLAKPTPKAETT-TKGP-ALTTPKEPTPTTPKKPRAPTTPKEPT 233
Db   129 TTTSTEPPSSS----TSPVQTTTTTAPETTSTEAPSSSSTTPVQTTTTTAPETTSTEPPSSS 184

Qy   234 .PTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKS 293
Db   185 TSPVQTTTTTAPETTSTEPPSSSSTTPVQTTTTTAPE---TTSTEP-PSSSTTPVQTTTTT 240

Qy   294 APTTPKEPAPTTPKKPAPTTPKEPTPTTPKEPAPTTTPKEPAPTTPKEPAPTTPKEPAPTAP 353
Db   241 APETTSTESPSSSSTTPVQTTTTTAPETTSTEP-PSSSTTPVQTTTTTAPETTSTEPPSSS 299

Qy   354 KKPAPTTPKEPAPTTPKEP--------------APTTTKEPSPTTPKEPAPTTTKSAPTT 399
Db   300 TTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSSTTPVQTTTTTAPET 359

Qy   400 TK--EPAPTTTKKSAPTTPKEPSPPTTKEP-APTTPKEPAPTTTPKKPAPKELAPTTTKGPTTSDKP-APTT 456
Db   360 TRTEPPSSSSTTPVQNTTTTTAPETTSTEPPSSSSTTPVQTTTTT-----APETTSTEPPSSST 415

Qy   457 EPAPTTTKKPAPTAPKEP--APTTKEPTAPTTPKKLTPTTPEKLAPTTPKEPAPTTPEEL 514
Db   416 TPVQTTTTTAPETTSTEPPSSSSTTPVQTTTTTTAPETTSTEP---PSSSTTPVQTTTTTA 471

Qy   515 APTTPEEPTPTTPEEPAPTTPKAAAPNTPKEP--APTTPKEPAPTTPKEPAPTTPKETAP 572
Db   472 PETTSTEP-PSSSTTPVQTTTTTTAPETTSTEPPSSSSTTPVQTTTTTAPETTSTESPSSS 529

Qy   573 TTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTSDKP-APTTPKGTAPTTPKEP 631
Db   530 TTPVQTTTTTAPETTSTEPPSSSSTTPVQTTTTTAPETTSTEPPSSSTTPVQTTTTTAPE- 588

Qy   632 APTTPKEP--APTTPKGTAPTTLKEPAPTTTPKKPAPKELAPTTTKGPTTSTPVSDKP-APTT 688
Db   589 --TTSTEPPSSSTTPVQTTTTTAPETTSTEPPSSSSTTPVQTTTTTAPETTSTEPPSSSTT 646

Qy   689 PKETAPTTTPKEPAPTTPKK---PAPTTPETPEPPTTS-------EVSTPTTTKEPTTI 735
Db   647 PVQTTTTTTAPETTSTEPPSSSSTTPVQTTTTTAPETTSTEPPSSSNTPVQTTTTTAPETTS 706

Qy   736 HKSPDESTPELSAEPT--PKALENSPKEPGVPTTKTPAAATKPEMTTTAKDKTTTERDLRTT 793
Db   707 TEPPSSSTSPVQTTTTTAPETTSTEPPSSSSTTPVQTTTTTAPETTSTEPPSSSNTPVQTTTTTAPETTS 766

Qy   794 -----PETTTAAPKMTKETATTTTEKTTTESKITATTTQVTTSTTTQDTTPFKITTTLKTTTLAP 849
Db   767 TTTTAPETTSTEPPSSSSTTPVQTTTTTAPETTSTEPPSSSSTTPVQTTTTTAPETTSTEP 824

Qy   850 KVTTTKTTITTTTEIMNKPEETA-KPKDRAINSKATTPKPQKPTRAPKKPTSTKKPKTMPR 908
Db   825 PSSSNTPVQTT---TTAPETTSTEPPSSSSTTPVQTTTTTAPETTSTEPPSSSSTTPVQTTT 882

Qy   909 VRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEG 968
Db   883 TTAPETTSTEPPSSSSTTPVQTTTTTAPET---TSTEP---PSSSTTPVQTTTTTAPETTS 936

Qy   969 -ETPHMLLRPHVFMPEVTP--DMDYLPRV------------PNQGIIINPMLSDETN---- 1010
Db   937 TEPPSSSSTTPVQTTTTTTVPCDCSLSYIDRVVYPFTEWWENKRDIIIQSYDSPRRTAFSF 996

Qy  1011 -----ICNGKPVDGLTTLRNGTLVAPRGHYFWMLS-------PFSPPSPARR 1050
Db   997 VSTPDIGTCTAILCIYTSEGISNL-NATL-----FIGLSDGSSIDLPF------- 1040

Qy  1051 ITEVWGIPSPIDTVPTRCNCEGKTFFFFKDSQYWRFTN 1087
Db  1041 ---YNVPTGLE-IWPEINCEGNK------WSYNN 1064
```

```
RESULT  6
Q9I7S1
ID   Q9I7S1       PRELIMINARY;   PRT;   1049 AA.
AC   Q9I7S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG5228 PROTEIN.
GN   CG5228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
```

```
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003495; AAG22353.1; -.
DR   FlyBase; FBgn0030561; CG5228.
SQ   SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;
```

Query Match 13.7%; Score 936.5; DB 5; Length 1049;
Best Local Similarity 34.6%; Pred. No. 6.1e-55;
Matches 371; Conservative 45; Mismatches 439; Indels 217; Gaps 54;

```
Qy   67 KDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTS----TTQHNKVSTSPKITTAKP 123
Db   30 KPNEIGTTAKPTTLKP---------------------------TEGTSAKPTTLKP       71

Qy  124 -----INPRPSLPPNSDTSKETPSL-----TVNKETTVERKETTTTINKQT--STDGKEKTT 171
Db   72 TEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEG----T   127

Qy  172 SAKET--QSIEKTSAK--DLAPTSKVLAKPT------PKAETTIKGPALLTTPKEPTPTT    220
Db  128 SAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLK              187

Qy  221 PKE-------PASTTPKEPT---PTTIKSAPTTPKEPAPTTPKEPAPTTKSAPTTPKE     272
Db  188 PTEGTSAKPTTLKPTEGTSAKPTTLK--PTEGTSAKPTTLKPTEG               243

Qy  273 TTPKEPAPTTPKEPAPTTTKSAPTTPKE--PAPTTPKKPAPTTPKEPAPTTPKEPT---P 327
Db  244 TTPAK---PTTLKPTEGTTAKPTTLNPTEGTSAKPTTLKPTEGTSAKP            300

Qy  328 TTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPK----EPAPTTTKEPSP  382
Db  301 TTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPFTDG            360

Qy  383 TTPKEPAPTT------TKSAPTTKEPAPTTKSA---PTTPKEPSPTTTKEPAPTTPKE   433
Db  361 TTAK---PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTTAK-             416

Qy  434 PAPTTPKEPAPTTPKEPAPTTPKETAPTT-----PTAPKEPAPTTPKETAPTT---PK   488
Db  417 --PTTLKPTEGTTAK--PTTLK-PTEGTSAKPTTLKPTEGTSAKPT              470

Qy  489 KLTPTI----TPEKLAPTTPKEPAPTTPEELAPTT--TPEEPTPTTPEEPAPTTPKAAAPN 541
Db  471 TLKPTEGTTAKPTTLKPTKGTSAKPTTLKPTEGTTAKPTTLKPT                530

Qy  542 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTA------------PTTLKEPAPTTP  591
Db  531 KGTTAKPTTLK---PTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTTA             587

Qy  592 K-----KPA-----PKELAPT--------TTKEPTSTTSDKPA--------PTTPKGT   623
```

RESULT 7
```
O76894
ID   O76894        PRELIMINARY;    PRT;    1795 AA.
AC   O76894;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:56G7.1 PROTEIN.
GN   EG:56G7.1 OR CG14796.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
Db  588 KPTTLKPTEGTSAKQPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTSAQPTTLKPT 647
Qy  624 APTTPKEPAPTTPKEPAPTTPKGTAPTTLKPEPAPTTPK----KPAPKELAPTTTKGPTST 679
Db  648 EGTTAK---PTTLK---PTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLNPTEG  701
Qy  680 TSDKPAPTTPKE--TAPTTPKEPAPTTPKKPA--PTTPEIPPPTTSEVSTPTT-TTKEPTT 733
Db  702 TSAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGTRTTAKPTT 761
Qy  734 TIIHKSPDESTPELSAEPTP------------KALENSPKEPGV--PTTKTT--PAATK    774
Db  762 TL--KPTEGT--SAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTERTSAQPTTLK   816
Qy  775 PEWTTTAKDKTTERDLRTT--------PETTTTAAPKMTKETATT---------TEKTTES 817
Db  817 PTEGTTAKPTTLKPTEGTSAQPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTTAK  876
Qy  818 KITATTTQVTSTTTQDTTPFKITTTLKTTTLAPKVTTTKK--TTTTTTEIMNKPEETAKPRD 875
Db  877 PTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTDGTTAKPTTLKPTEGTSAKPTTLKPTE  936
Qy  876 RATNSKATTPKPQKPRPTKAPK---KPT---STKKPKTMPRVRKPKTTPT---PRKMTSTMP- 926
Db  937 -GTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEGPSAKPTTLKPTERTSAQPT  995
Qy  927 ELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEG-ETPHMLLRP         977
Db  996 TLKPTERTSAQ--PTTLKPTEGTTAKPTTLKPTPKGTSGQANNFETKRKERRP        1045
```

```
RA   Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RA   Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.;
RN   [3]
RP   SEQUENCE FROM N.A.
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003421; AAF45644.1; -.
DR   EMBL; AL031028; CAA19845.2; -.
DR   FlyBase; FBgn0025390; EG:56G7.1.
DR   InterPro; IPR002557; Chitin_binding.
DR   InterPro; IPR002965; Chitin_extensn.
DR   Pfam; PF01607; Chitin_bind_2; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00494; ChtBD2; 2.
SQ   SEQUENCE 1795 AA; 194464 MW; 07F10C129BD95557B CRC64;

Query Match     12.7%;     Score 865.5;   DB 5;    Length 1795;
Best Local Similarity   28.5%;   Pred. No. 6.4e-50;
Matches  276;   Conservative 129;   Mismatches 447;   Indels 115;   Gaps 28;

Qy   71 KNRTKKKPPTPKPPVDEAGSCLDNGDFKVTTPDTSTTQHNKVSTSSPKITTAKPINPRSL 130
           |||
Db  431 KNDVTEAPEIKSPL------------KGLHLSENIVLLPETTTT--TTTTTKPVVLICPISPPDTT 483

Qy  131 PPNSDT---------------SKETSLIVNKETTVETKETTTINKQTSTDGKEKTTSAKETQSI 179
           | | |
Db  484 PKPSTTTAVTKSTPKISSTEQHSTTTAKTTTTKRPTTVTEKTSSATEKPRTTVVTTTQK 543

Qy  180 EKTSAKDLAPTSKVLAKPTPKAETTKGPALITPKEPPTTPTKEPASTTPKEPPTTIKS 239
           |:|              |||           |||      |||
Db  544 RSTTTHNTSPDTIKTTIRSTTLSPKTTTPSTTPSTTPST----TTPSTTPST--- 594

Qy  240 APTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKSAPTTP 298
           |||           |||           |||
Db  595 --TTPSTTTPPSTTTVKVSTHRPRTTSQKTTVASTTVKK------TTTSPKTTKTTDIPTST 648

Qy  299 KEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTKEPAPTTPKEPAPTTAPKKPAP 358
           |:|           |:|
Db  649 TSKLSTTTQKTTTTHKFTAATTISTEKPKTTTEKTSTVSTTVTTKKSTESSSPKPTSSTGKPT 708

Qy  359 TTPKEPAPTTPKEPAPTTTKEPSPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEP 418
           |||           |||           |||
Db  709 TTPKPSTRTPTTPTTKVTTTQITTTTPLRSSTETTSTQPPTTTTPQPTTTTTLTVTPKTS 768

Qy  419 SPTTTTKEPAPTTPKEPAPTTPK-----------KPAPTTPKEPAP---------------TTPKE 457
           |||           |||            |
Db  769 TTTTTTEKPITSSPKPTTTOKTTSTAPNTTKVAITTQKETTPTTQSTTTSTITTFTRKTTTNN 828

Qy  458 PAPTTTKKPAPTAPKEPAPTTPK-ETAPTTPKKLTPTTPEKLAPTTPEKP--------TTPKE 508
           |:|:|:||           |:|
Db  829 PEPTSTEKPITSTTPKPSTTSTPPKTTSTVASSTEKTTISSPKPTTEKSTENPTTNSVKTSAL 888

Qy  509 TTPEELAPTTPEEPTPTTPEEPAPTTPKEPAPTTPKAAAPENTPKEPAPTTPKEPAPT--TPKEPAPTT 566
           :|| | |     |   |||           |||
```

```
Db  889 TSSTQORATSTTSEPTKTT-QNITTTTPKPTTLKTSTQEATTSTQKVSTVTITTKAPESS 947
Qy  567 PKETAPTTPKGTAPTTLKEPAPTTPKPAPKELAPTTTKEPT--STTSDKPAPTTPKGTA 624
Db  948 PLTTLSTEEPNTTPKPLRTTTPPTTTISVTTATTRITTTTISESSTETTSTQKPKSTTPTSTT 1007
Qy  625 PTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP-TTTKGPTSTTSDK 683
Db 1008 RTTPKVTTVIVSTQNPTTTSKTSTVTI----TTP-NPSPSTQRPTTTTRQPTISTTAST 1061
Qy  684 PAPTTPKETAPTTTPKEPAPTTPKKPAPTTPETPPTTPTTSEVSTPTT-TKE-------- 731
Db 1062 TSIGTTRIPTTTNPQNSTSSSTDLTTVTRPCPDPDSTSDKNTNTACTQELQQVNLELQS 1121
Qy  732 ----------PTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAATKPEMTTTAKD 783
Db 1122 PQKQEQFTHTRTHTALTGSRNTLGGQEVPDYMDDAPSSAEAESGQATTAKAPTMSTLAAA 1181
Qy  784 KTTER---DLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKIT 840
Db 1182 HLLQKLFHIISTTPPSREHAP--TQRPSSQPSSSQRSR-GVTIAQMARHNLATSKPFIAH 1238
Qy  841 TLKTT----TLAPKVTTTTKKPITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKA-PK 895
Db 1239 SIRLSTIQQLASTQKRSIPPPKTLVHTNHTTKEPED-SEYYDSETSEGYTDEDNEVLDKTQPR 1297
Qy  896 KPTISTRKPKIMPVRKPKTTPTPPRKMTSTMPELNPTSRIAEAMLQTTTRRPNQTPNSKLVE 955
Db 1298 AMSSTTVAAVLPAVPSTTTEREPOK-TSSSP--SPT----KATSSTTTQPIETTTGDLEY 1350
Qy  956 VNPKSED 962
Db 1351 DSSGSSD 1357

RESULT   8
O96449
AC   O96449;
ID   O96449     PRELIMINARY;   PRT;   1489 AA.
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN   CAR90.
OS   Phytophthora infestans (Potato late blight fungus).
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RACE 1-11;
RA   Goernhardt B.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061185; AAC72308.1; -.
SQ   SEQUENCE 1489 AA; 764C779D1C2F5163 CRC64;

Query Match     12.5%;     Score 855;   DB 10;    Length 1489;
Best Local Similarity   31.8%;   Pred. No. 2.7e-49;
Matches  362;   Conservative 52;   Mismatches 496;   Indels 228;   Gaps 51;

Qy   37 FERGRECDCDAQCKKYDKCCPDYESFCAEVKDN--KKNRIKKKPPPKPPVDEAGSGLD 93
           :|    :|     :|       ||   :|
Db  301 YEPSDETEAPTEGTTV---VPREETTAAPSEDTTYAPREVTPYAPTEKPVDVEETTVTE 357
Qy   94 NGDFKVTTPDTSTT------QHNKVSTSPKITTAKPINPRSLP-----------P 132
           :|                 :|
Db  358 ESTYAPTPKSETNAPTERMHYAHIEKPCDTEVTMYAPTEETTYAPTEETTYAPTEETTYAP 417
Qy  133 NSDT----SKETSLIVNKETTVETKETTT---TNKQTSTDGKEKTTSAKETQSIEKTSAK 185
           :|       :|||                 :|
Db  418 TEEETPYEPTEETTYTPTEETTYAPTEETTYAPTEKTTYAPTEETPYEPTEET 477
```

```
RESULT  9

Q9SPM0
ID   Q9SPM0       PRELIMINARY;   PRT;   1315 AA.
AC   Q9SPM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN.
GN   PEX2.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=POLLEN;
RA   Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA   Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT   "Pollen Extension-like (Pex) Genes in a Monocot and a Dicot.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF159297; AAD55980.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR001998; LRR.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR   SMART; SM00370; LRR; 3.
SQ   SEQUENCE   1315 AA;  134401 MW;  64C97A2A01F0936F CRC64;

Query Match             11.9%;  Score 811.5;  DB 10;  Length 1315;
Best Local Similarity   24.6%;  Pred. No. 2e-46;
Matches   262;  Conservative   89;  Mismatches  473;  Indels  243;  Gaps  35;
```

Left column alignment:

```
Qy   186  DLAPTSKVLAKPTPKAETTTKGPALTTPKEPPTTPKEPPTTPKEPASTTPKEPPTTIKSAPTTPK  245
Db   478  TYAPTKETTYAPT---EETTYASTEETTYAPTEETTYAPAEETPYEPTEET-TYAPTEET  533

Qy   246  EPAPT-TTKSAPTTPKEPAPT--TTKEPAPTTPKEPAPTTTPKSAPT------  296
Db   534  TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETMYA  592

Qy   297  ----TPKEPAPTTPKKPAPTTPKEPAPTTPKEPTTTPKEP------------AP---  335
Db   593  PIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYASTEETTYAPTEE  652

Qy   336  TTKEPAPTTPKEP-----APTAPKKPAPTTPKEPAPTTPKEPAPT--TTKEPSPTTPKE  387
Db   653  TTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYE  712

Qy   388  P------APT-TTKSAPT-------------TTKEP-----APT-TTKSAPT--TPKEPSPT  421
Db   713  PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYE  772

Qy   422  TTKEP-------APTTPKEPAPTTPKKPAPTTPKEPAPT--TPKEPAPTTTKKPAPTAPKE  473
Db   773  TTYEPTGEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYE  832

Qy   474  PAPTTPKETAPTTPKLTPTTPEKLAPTTPEKPAPTTPEELAPT--TPEEPTPTTPEEP-  530
Db   833  PTEET--TYTPTEETTYAPTEETTYAPTEETTYAPTEETTYAPT  890

Qy   531  -------APT-------TPKAAAPNTPKEPAPTTPKEP---------APTTPKEPAPTTPKE  569
Db   891  KETTYAPTEETTYASTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPAEETPYE  950

Qy   570  TAPTTPKGTAPT--TLKEPAPTTPKEPAPK-ELAPT--TTKEPTSTTSDKPAPTTPKGTA  624
Db   951  YAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPIEETTYAPT  1010

Qy   625  PTTPKEPAPTTPKEP--------APTTPKGTAPT---------TLKEPAPTTP  660
Db  1011  EETTYAPAEETPYEPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETP  1070

Qy   661  KRP------APKE---LAPT--TTKGPTSTTSDKPAPTTPKETAPTTPKKPA  709
Db  1071  YEPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPT  1130

Qy   710  PTTPETTPPTTPPTTSEVSTPTTTKEPT--TIHKSPDESTPELSAEPTPKA-LENSPKEPG---  763
Db  1131  EETTYAPIEETTYAPTEETTYGPTEETTY-APTEATTYAPTEETTYGPTGET  1189

Qy   764  --VPTTKIPAATKPEMTTTAKDKTERDLRTTP----ETTTAAPKMTKETATTTEKTTES  817
Db  1190  TYAPTEETTYAPTEETTYAPTEETTYEPAEEESTSTVSTEKPCNTEEF  1244

Qy   818  KITATTTQVTSTTTQODTT---------PFKITT----LKTTTLAPKVTTTPKKITTTEIM  864
Db  1245  TTYAPTEETTYAPTEETTYAPIDEETTYGPTEETTYAPTEETTYA  1304

Qy   865  NKPEETAKPKDRAT--NSKATTPKP-QKPTKAPKKPT--------STKKP-KIMPR  908
Db  1305  PTEETTYEPTGETTYAPTEETTYAPTEETTYAPAEEESTSTVSTEKPCNTEEF  1364

Qy   909  VRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP---NQTPNSKLVEVNPKSEDAG  964
Db  1365  TDEPTDEPTDE--PSDEPTDEPTDLPTDEPSTPCDNQGINGIGVENKVRYNNAG  1420
```

Right column alignment:

```
Qy    32  RCFESFERGRRECDCDDAQCCKKYDKCCPDYESFCAEVKDNKKNRIKKKPTP----K  81
Db   390  QCAPVLARPVEDVDCSKH------VCAGY--------PTPGGPPSSSVPGK  425

Qy    82  PPVVDEAGSGLDNGDFKVTTPDTSTTQOHNKVSTSPKITTAKPINPRPSLPP-----  132
Db   426  PPSVPG----------KPAAPAPMPTPHTPPDVSP-----EPL-PEPSVPAPAPMRMPT  469

Qy   133  -NSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTS  191
Db   470  LRSPPADEYIPT----PPVPAKSPPGTSPPASRGAPPLQOQPPAASSPPATPVKSSPPPA  525

Qy   192  KV-----------LAKPTPKAEFTTKGPALTTPKEPPTTPKEPTTTPTTI  237
Db   526  AVVLPPPAKTPSPPAPVASPPEEAPVSSPQPQVKSPSPPPAPVASPPPPWKSP--PPPARV  583

Qy   238  KSAPTTPKEPAPTTTKSAPTTP-KEPAP------TTTKEPAPTTPKEPAPTTTKEPAPT  289
Db   584  ASPPPLMKSPPPAPVASPPDQFLKSPPPPVLWLSTPSVKSPPPPVPVASPPPPVPVKSPPPL  643

Qy   290  TTKSAPTTPKE---PAP--TTP----KKPAPTTPKEPAPTTPKE-PTPTTPKEPAPTTK  338
Db   644  APVSSPPSPPVKLPPLPAPGKSPPPEEEKPTPTPVKSSPPPPLPPPPLTTSPPPQEK  703

Qy   339  EPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKLLTPTTPKEKLAPTTPTKPKEPAPTTTKKSAPT  398
Db   704  PTPPSTPSKPPPPSPVEELPPPSKSSPPPEEPVSSPPQOAPKSSSP----PAPVS-SPPP  756

Qy   399  TTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTTKEPAPTTPKEPAPTTPKEP  458
Db   757  LKSSPPPPVPESSPPPTPKSSSPPPAHVSSSPPEAEKSSSPPLAPISSPPAPVSS  816

Qy   459  APTTTKKPAPTAPKEPAPTTPKETAPTTPKLLTPTTPKEKLAPTTPTTPKEPAPTTPEELAPTT  518
Db   817  PPQVEKTSPPPAPVSSPPPTPKSSSPPLAPVSSPPQVEKTSPPPAPVSSPPPLEPKPSSPPS  876

Qy   519  PEEPTPTTPEEPAPTTPKKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGT  578
Db   877  SVSSPPTTVKSSPPPAPLSSPPPMTPKSSSPPAHVSSSPPEAEKSSSPPLAPISSPPSEPKSP  936

Qy   579  A-------------PTTLKEPAPTTPKPAPKELAPTTTKEPTSTTSDKPAPT-----TPKG  622
Db   937  SSPPMVEKTSPPPATVSSPPPTPKSSSPPPL---APVSSPPPVKSSPPPAPVSSPPPTPKP  993
```

```
Qy   623 TAPTTPP----------KEPAPTTPKEPAPTTPK-GTAPTTLKEPAPTTPKKPAPKELAPTT 672
             |||                 ||| | ||      :|||: |||||||||  :|
Db   994 LPPPAPVSSPPPVVKSSPPPTPVSSPPPPKPLPPPPTPVSSPPPTXKPLPPP----APVS 1050

Qy   673 TKGPTPSTTSDKPAPTT---PKETAPTTPK-----EPAPTTPKKPAPTTPETPPPPTTSEVSTP 726
             ||||           |||     |||        |||    |||  :||
Db   727 TTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTT 786

Qy   787 ERDLRTTP---------ETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITT 841
             ::||           |||||||||||||||||||||||||||||||||||||||||||
Db   842 LKTTTLAPKVTTTKKTITTTTEIMNKPEETAKPKDRATNSKATTPKQKPTKAPKKPTSTK 901

Qy   902 KPKTMPRVRKPKITTPTRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSE 961
Db  1051 SPPPAP--VKPPSLPPPAPVSSPPPAV------------------TSAP----------PKKE 1110

Qy  1051 SPPPVVKSSPPPAVSLPPPTKRPSPPRTRVSSPRPVVKCCPPPTLVSSPPPAPKSLPPP 1110

Qy  1111 TPVSSP-------PPEVKSPPPTPVSSPPPPAPKSSPPPPTPPVSSPPP-------- 1150

Qy  1151 --ELKSSPPPAPVSSPPSPSAPKSSPPPPEVKSSPPPPAPISSPPPPAKSP----; 1203

Qy  1204 -----PPPA------PMSSLPPPVKSPPPPAVSSPPPPMKSPPPPAPIS 1242

Qy   962 DAGGAEGETPHMLLRPHVFMPPEVTPDMDYLPRVPNQGIIINPMLSDE 1008

Qy  1243 SPPPAP--VKPPSLPPPAPVSSPPPAV----------TSAP------PKKE 1275

Qy  1276 EDSTA---------PPAEALPPPSENDILLPPIMANK 1303
```

RESULT 10
Q20007

```
ID   Q20007;       PRELIMINARY;       PRT;   1274 AA.
AC   Q20007;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   COSMID F35A5.
GN   F35A5.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Leimbach D.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U46675; AAB52641.1; -.
SQ   SEQUENCE   1274 AA;   138065 MW;   07F6B0D292C1799F CRC64;
```

Query Match   11.8%;   Score 802;   DB 5;   Length 1274;
Best Local Similarity   28.0%;   Pred. No. 8.2e-46;
Matches   290;   Conservative   102;   Mismatches   414;   Indels   230;   Gaps   56;

```
Qy    71 KNRTKK-KP------TP-----KPPVDEAGSGLDNGD----FKVTTPDTSTTQHNKVSTS 115
Db   274 KNPTKKWKPPWEDDETPVEEVKPPVPEKKAPVLKKKDPAPAAKARDPSPSKAAPKKVEPS 333

Qy   116 PKITTTAKPI-NPRPSLPPNSDTPSKEISLTVNKETTVETKETTTNKQTSTDGKEKTTSAK 174
Db   334 SPVVPPPPVKNPVKKYKPPWEVDDEPAEEVKKPSAPEKKTPVLKRKEPEPSSTTPSSDPS 393

Qy   175 ETQSIEKTSAKDLAPTSKVLAKPTPKAE------TTTKGPA------------LTTP 213
Db   394 PKKAAPAVKRDSSPKKATPLQADPKRAQEVPPTPVKNPVKKYKPPWEVDDEDPVEEVKQP 453

Qy   214 KEPTPTTP-----KEPASTTPKEP----TP-TTIKSAPTTPKEPAPTTTKSAPTTPKEPAP 264
Db   454 EAPAKKTPVLKRKEPAAKDTAKPAISKTPETPEKKDPVKPRDSSPKKVAAKPDSAQAPA- 512

Qy   265 TTTKEPA------PTTTKEPAPTTTKEPAPT-------TTKSAPTTP 298
Db   513 TPVKNPVKKWRPPWEDDETPADDVSKPTDAKKTPSLAKKDPAPAKESLRPKADTKAPAKP 572

Qy   299 KEP--------APTTTPKEPAPTTPKE-----PAPTTPKEPTPTTPKEPAPTTTKPKEP-EP 348
Db   573 RDPSPKKVAPTPAPEKKTPVLAKKEPAGPADSKTPEPEKSKPRDPSPKKAVPAKVPKTEV 632

Qy   349 APTAPKKPAP-----------TTPKEPAPTTTPKEPAPTTTKEP-----SPTTPKEPAPTTTKSAPT 398
Db   633 APAAVKPEEPISKPKDTAPKKAEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPE 691

Qy   399 TTTKEPAPTTTKSAPTTP-KEPSPTTTKEPAPTTPK--EPAPTTPKKPAP--TTPKEPAPT 453
Db   692 PEKK-TPVLAKKAPTKPDSEAAADPVSGPSSKDPKLAKKAPVKPRDPSPMKAVPIKPAPK 750

Qy   454 TPKEPAPTTTKKPAPTA-PKEPAPTTPKETAPTTPKKLPTTTP------EKLAPTTPEKPA 507
Db   751 T--EVPPAVVKKPEPVAKSRDPSPKKAK-AEPNSP--VVPPTPVKNPVKKWKPPWEDDDA 805

Qy   508 PTTTPEELAPTTPEEP-----TPTTPEEPAPTTPKKAAAPNTPKEPAPTTPKEPAPTT-PKE 561
Db   806 PAEPVNVPEPEKKTPVLAKKTPVKPRDPSPKKAVPAKPSTKTDAPPVSVKKPEPVSKPKE 865

Qy   562 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPT--TTKEPTSTTSDKPAPTTP 620
Db   866 PSPKKAEPNSPVVP----PTPVKNPVKKW-KPPWEDDDEPTEEVKKPESE--PEKKTPVLA 918

Qy   621 KGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPT--------TPKKPAP-----KEL 668
Db   919 K-KEPEKPRD-APKVAAKPRDSPKKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEA 974

Qy   669 APTTTKGPTS-----TTSDKP-----APTTPKET------APTTPKEPAPTTTPKPAPT 711
Db   975 PPTPVKNPVKKWKKPPWEDDDEPAEPVSAPEPEKKTPVLAKKAPAKRDP---SPKKAAPV 1031

Qy   712 TPETPPPPTTSEVSTPTTTKEPTTIHKSP-----DESTPELSA-EP----TPKALENSPKEPG 763
Db  1032 AAK-PDPKIPEV-PPTPVKNPVKKWKPPWEDDDEPSEPVSAPEPEKKTPVLAKKAPTKPA 1089

Qy   764 V------------------KPEMTTTAKDKTTERDLRTTPETT--TAAPK 802
Db  1090 TKPDSEAAADPVSGPTSKDPKLSKKAPVEKRKPTTPDKDDKLKPSPAKKPEKAPEEPAAPK 1149

Qy   803 MTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTTE 862
Db  1150 KKKPVWDDDPDPDPEAADTVPAPSKKKPDTEDPDPLG---------GPKTKDPK------ 1193
```

```
Qy   863  IMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRKPK--------- 913
          ||:|      |.||| :|      |:||    |  |:|   ||  |::|
Db  1194 -LNKKAPAEKPTEK-------PKPKEVSKEPEKPTEPRKP-AAPKKWKPPWEDDPDEPE 1243

Qy   914 ---TTPTPRKMTSTMP 926
         |.|:|  |||
Db  1244 ADFTMPAPKKPDTEDP 1259
```

RESULT 11

Q9LIE8   PRELIMINARY;   PRT;   1480 AA.

```
ID   Q9LIE8
AC   Q9LIE8;
DT   01-OCT-2000 (TREMBLrel. 15, Created)
DT   01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP001306; BAB03062.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE  1480 AA;  147153 MW;  D1AC0C79F155E732 CRC64;
```

```
Query Match          11.7%;        Score 800.5;  DB 10;  Length 1480;
Best Local Similarity  29.6%;        Pred. No. 1.2e-45;
Matches  307;  Conservative  54;  Mismatches  505;  Indels  171;  Gaps  45;
```

```
Qy    75 KKKPTPKPPVVDEAGSGLDNGDFKVTTPDTISTTQHNKVSTS----PKITTAKPINPRPSL 130
Db   117 KRPHPKPPNVK------------PHPHPKPRTKPHPHPKPPTKHNPHPKPPTIKPPKPPSV 167

Qy   131 PPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTISAKDLAPT 190
Db   168 KPPPSTPK------------PPTTNPPPSTPQPPTHKPPPCTPPPPVASPPMATPP 211

Qy   191 SKV--LAKPTPKAETTKGPALTTPKEPTP--TTP--KEPASTTPKEPTPTTIKSAPTTP 244
Db   212 TQMPIATP-PIAKSPVATPPIATPPIATPPIITPPVATPPIITPPIANPPIIMPPIATP 270

Qy   245 KEPAPTTT----------KSAPTTPKEPAPTTT----------KEPAPTTPKEPAPTTTE 285
Db   271 PVAAPPITNPPISKPPVTTP----PTTTPPIAKPPIATPPISTPPAATPPIITLP 326

Qy   286 PA-------PTTTKSAPTTP-KEPAPTTPKEPAP--TTPKEPAPTTPKEPTPTTPKEPAP 335
Db   327 PAKPPVAISPIVT--PPVTPIAQPPVATPPIATPPIATPPIATPPISKSPPISPPISESP 384

Qy   336 TTKEPAPTTP-KEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTT----KEPSPTTPKEPAP 390
Db   385 VATPPTPATSPIKTPPPAKPPVATPPIAKSPIATPPTATPPVATPPIEKPPVATPPTTPP 444

Qy   391 TTTKSAPTTKEPAPTTTKSAPTTKEPAPTTPKEPAPTTPKEPAP---SPTTTKEPAPTTP 445
Db   445 PPVAKPPVETPPIATPPTAKPPISTPPPISKPPVATPPAATPPITTPTPVKRPVA 501
```

```
Qy   446 TPKEPAPTTPKEP--APTTTKKPAPTAP--KEPAPTTPKETAPTT-------PKKLTP--T 493
Db   502 TPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVATPPIIAKPPVVTPPTT 561

Qy   494 TPEKLAPTTPEKPAPTTPPEELAPTT--TPPKEPAPTTPEEPAPTTPKAAAPNT---PKEPAPT 549
Db   562 TPPIATPPVAKPPVATPPIATPPTAKPPISTPPISKSPVATPPAATPPITTPPAKPPVA 621

Qy   550 TPKEPAPTTPKEPAPTTPKETAP--TTPKGTAPTT---LKEPAPTTPKKPAPKELAPTT 604
Db   622 TPPIATPPIAKPPVATPPIATPPIATSPVAKPPVATPPIKTPPPAKPPVAIPPIATPPVA 681

Qy   605 KEPPSTTSDKPAPTTPKGTAP--TTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKP 663
Db   682 KPPVAT---PPTATPPIATPPIATPPVVTPPTATPPVATPPIAKPPVATPPII--PPTATPPVA 736

Qy   664 APKELAPTTTKGPTST--TSDKPAPTTPKEPAPTT-TPKEPAPTTPKKPAPTTP-----E 714
Db   737 MPPIATPPPTAKPPIATPPIAIPPVAKPPVVTPPTATPPIATPPIAKSPVATPPTATPPVA 796

Qy   715 TPPPTTSEVSTPPTTTKEPTTIH--KSPDESTPPLSAEPTPKALENSRKEPGVPTTKTPAA 772
Db   797 TPPIAKPPVAIPPTTAPPIATPPVAKPPVATPPIATPPTAKPPILTPPISKPPVATPPAA 856

Qy   773 TKPEMT-TTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKIIATTTQVTSTTT 831
Db   857 TPPIITPPPAKPPVATPPIATPP--IAKPPVATPPTATPPIATSPVAKPPVAIPPIKTPP 914

Qy   832 QDTTPFKIITILKTTILAPKVTTVTTKKT----IITTEIMNKPEEPAKPKDRATNSKATTPKP 887
Db   915 PAKPPVAIPPIATPPVAKPPVATPPTATPPIATSPIATPPVVT---PPTATSPVATPPIA 971

Qy   888 QKPTKAPKKPPTSTKKPKTMPRVRKPKTTPTPPRKMTTSTMPELNP-------TSRIAE--- 936
Db   972 KPPTTTP--PTAT-PPVAMPPIATP---PTAKPPVATPPIANPPVEKPPVATPPIAKPPI 1025

Qy   937 -----AMLQTTTRPNQTPNSKL--VEVNPRKSEDAGGGAEGETPHMLLRPHVFMPEVTPDMD 989
Db  1026 VLPPPIAKPPVETSPIATPPTATPPVAIPP-----------VVKPPVAIPPITKPPV 1070

Qy   990 YLPRVPNQGIIIINPMLS 1006
Db  1071 ATPPVTNPPTAMPPIVT 1087
```

RESULT 12

Q41805   PRELIMINARY;   PRT;   1188 AA.

```
ID   Q41805
AC   Q41805;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z34465; CAA84230.1; -.
DR   Mendel; 14346; Zeama;2368;14346.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 3.
DR   SMART; SM00370; LRR; 4.
KW   Signal.
FT   SIGNAL        1     27       POTENTIAL.
SQ   SEQUENCE  1188 AA;  120980 MW;  2C77C7F8D7130149 CRC64;
```

```
RESULT 13
ID   P70670      PRELIMINARY;      PRT;   2187 AA.
AC   P70670;
DT   01-FEB-1997 (TREMBLrel. 02, Created)
DT   01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE   MUSCLE-SPECIFIC FORM GP220).
GN   NACA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96312450; PubMed=8698236;
RA   Yotov W.V., St-Arnaud R.;
RT   "Differential splicing-in of a proline-rich exon converts alphaNAC
RT   into a muscle-specific transcription factor.";
RL   Genes Dev. 10:1763-1772(1996).
DR   EMBL; U48364; AAB18734.1; -.
DR   EMBL; U48363; AAB18732.1; -.
DR   MGD; MGI:106095; Naca.
DR   InterPro; IPR002715; NAC.
DR   InterPro; IPR003037; TS-N.
DR   Pfam; PF01849; NAC; 1.
DR   Pfam; PF02094; TS-N; 1.
SQ   SEQUENCE   2187 AA;  220599 MW;  003646AA864DEBFD CRC64;
```

```
Query Match   11.5%;  Score 784;  DB 10;  Length 1188;
Best Local Similarity   29.9%;
Matches  230;  Conservative   59;  Mismatches  359;  Indels  122;  Gaps   33;

Qy    78 PPTPKPPPVVDEAGSGLDNGDFKVTTPDTSTQHNKVSTSPKITTAKPINPRPSL----PPN 133
            ||   :||  || |
Db   460 PTPHSPPAD------DYVPPTPPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQ 510

Qy   134 SDPSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV 193
Db   511 APVG---------SPPPPVKTTSPPAPIG---SPSPPPVSVV 541

Qy   194 ------LAKPTPKAETTTKGPALTTPKEPTTPTPKEPASTTPKEPTPTTPKEPA 248
Db   542 SPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVASPPPPAVKSP--PPPLVASPPPPVKSPP 599

Qy   249 PTTTKSAPTTP--KEPAPTTPKEPAPTTPKEPAPTTTKSAPTTP-KEPAPTT 305
Db   600 PPAPVASPPPPVKSPPPPTPVASPPPPAVASPPPPAVSPPPPPPVSSPPPPEKSPPPPP 659

Qy   306 PKKPAPTTPKEPAPTT------PKE---PTPTTPKEPAPTTPKEPAPTTPKEPAPTARKKP 356
Db   660 PAKSTPPPEEYPTPPTSVKSSPPPPTLIPSPPPQEKPTPPSTPSKP-PSSPEKP 718

Qy   357 APTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTIKSAPTTPK 416
Db   719 SP--PKEPVSSPPQTP----KSSPPPPAPVSSPPPTPVSSPPPALAPVSSPP------- 768

Qy   417 EPSPTTTPKEPAPTTPKEPAPTTTPKKPAPTTPKEPAPTTTKKPAPTTAPKEPAP 476
Db   769 PPAPLSSPPPPAPQVKSSPPPPVQVSSP-PPAPKSSPPPLAPVSSPPPPQVEKTSPPPPAPLSSPP 827

Qy   477 TPKETAP-----TTP----KKLTPTTPEKLAPTTPEKPAP----TTPEELAPTTPEEPT 523
Db   828 LAPKSSSPPHVVVSSPPPPVVKSSPPPPAPVSSPPPEVKPSTP--PA 885

Qy   524 PTTPEEPAPTTPKAAAPNTPKE-PAPTTPKEPAPTTPKETAPTTPKGTAPTT 582
Db   886 PTTVISP-PSEPKSSPPPTPVSLPPPIVKSSPPPAMVSSP-PMTPKSSPPPVVVSSPPPT 943

Qy   583 LKE---PA----PPTPKK----PAPKELAPTTTKE---PISTTSDKPAPTTPKGTAPTTP 628
Db   944 VKSSPPPPAPVSSPPATPKSSPPPAPVNLPPPEVKSSPPPTPVSSPPPA---PKSSPPPAP 1000

Qy   629 -KEPAPTTPKEPAPTTPKRGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTT 687
Db  1001 MSSPPPPEVKSSPPPPEVKSSPPPPAPVSSPPPPAPVSSP-----PPVKSPPPPAPVSSSPPPPV 1057

Qy   688 -TPKETAPTTPKEPAPTTPEETPPPTTSEVSTPTTTKEPTTIIHKSPDESTPEL 746
Db  1058 KSPPPPAPISSPPPPVKSPPPPAPVS---SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS 1115

Qy   747 SAEPTPKALENSPKEPGVPTTKTPAAATKPEMTTTIAKDKTTERDLRTTPET 796
Db  1116 SPPPAP-------VKPPSLP-PPAPVSSPPPVVTPAPPKEEQSLPPPAES 1158
```

```
Query Match   11.4%;  Score 776.5;  DB 11;  Length 2187;
Best Local Similarity   27.6%;
Matches  290;  Conservative  119;  Mismatches  416;  Indels  227;  Gaps   48;

Qy    57 PDYESFCAEVKDNKKNRTKKKPFKPPPVVDEAGSGLDNGDFKVTTPDTSTQHNK----- 111
Db   809 PKVDPIMSDVTPTSPKKTSATAVPK---------------DTSATLSLKSVPAV 847

Qy   112 VSTSPKITTAKPINPRPSLPPNSDDTSKETSL--TVNKET-------TVETKETTTTNKQTS 163
Db   848 TSLSPKPAPVAPSNEATIVPTEIPTSLKNALAAAATPKETLATSIPKVTTSPSQKTPKSVS 907

Qy   164 TDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKE 223
Db   908 LKGAPAMISKKAT---EIAASKDVSPSQ--FPKEVPLLQHV---PPTSPRKSPVSDTLSG 959

Qy   224 PASTTPKEPTPTTIKSAPTTPKE-PAPTTTKSAPTTPK--------------EPAPTTTKEPAPT 273
Db   960 ALTSPPPKGPPATLAETPTYPKKSKPKPAASKKTPATPSPEGVTAVPLEIPCSKKAPKTA 1019

Qy   274 TPKEPAPTTTKEPAPTT---TKSAPTT------------------TPKEPAPTTPKK-P 309
Db  1020 APKESSSATSSSKRAPKTAVSKEIPSKGVTAVPLEISLPLKETSKSATPGEKSASSPKRSP 1079

Qy   310 APTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPATAPKKPAPTT--PKEPAP- 366
Db  1080 KTAGPKE------TPPGGVTAVPPEISLPPKETPQNATPNESLAASSQKRSPKITSVPKETPPG 1136

Qy   367 ---TTPKEPAPTTTKEPSPTTPK--------------EPAPTTTKSAPTTTKEPAPTTT-KSAPTTPK- 416
Db  1137 GVTAMPLEIPESAPQKAPKTAVPKQIPTPEDAVTILAGSPLSPKKASKTAAPKEAPATPSV 1196

Qy   417 ------EPSPTTTPKEPAPTTPKEPAPTTTPKKPAPTT---PKE---------PAPTTP 455
Db  1197 GVIAVSGEISPSPKKITSKTAAPKENSATLPPKRSRKTAAPKETPATSSEGVTAVPSEISP 1256

Qy   456 KEPAP-----------TTTKPAPTTAPKEPAPTTPKETAPTTPKKLITPTTPEKLAPTTPK-PA 507
Db  1257 SPPTPASKGVPVTLPKGAPNALAE-SPASPKKVPVKTAAPEETSTTP------SPQKIPK 1309

Qy   508 PTTPEELAPTTPEEPTTTT---------PEEPAPTTPKAAAPNTPKEPAPTTPKE-PAPTT 558
Db  1310 VAGPKEASATPPSKKTPKTAVPKETSAPSEGVTAVPLEIPPSPRKAPKTAAPKETPAPS- 1368

Qy   559 PKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTP------------KKPAPKE 598
Db  1369 -PEGATTAPVQIPPSPRKGSKKAGSKE-TPTTPSPEGVTAAPLEIPISSKKTSKMASPKE 1426

Qy   599 LAPTTPKEPTS------TTSDKPAPTTPKGTAPTTPKE-PAPTTPKGTAPTT 651
Db  1427 TLVTPSSKKLSQTVGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLITPSPK-DAPTT 1485

Qy   652 LKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKGPTSTTPKEPAPT 711
Db  1486 LAE-SPSSPKK-APKTAAPPSER-VTTVPPEKPA-TPQKASGTTASKVPVPAETQEVAVS 1541
```

```
Qy    712  TPEP-----PPTTSEVSTPTTK------EPTTIHKSPDESTPELSAEPTPKALENSPK  760
                ||||      ||::|| :|       |:||:    |:||:|     |:||    |||
Db   1542  SREPTPVTPAVPPVVNPSSHKKTSKTIELKEAPAILPPSPTKSPKIPSSKKAPRT--SAPK  1599

Qy    761  E-PGVPTIKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKEFATTTEKTTESKI  819
              | |  | | |   |: ||        |  :|| :|  |      |    |||:| ||
Db   1600  EFPASPSIK-PVTT--SLAQTAPSLQKAPSTTIPKENLAAPAV----LPVSSKSPAAPA  1652

Qy    820  TATTTQVTSTTTQDTTPFKITTLKKTTLLAPKVTTTTKKTITTTEIMNKPEETAKPKPDRATN  879
              ||   :    |  |||  :|||    |||||| | | ||::|||  |       ||:
Db   1653  RASASLSPATAAPQTAPKEATTIPSCKKAAATETPIEISTAPSLEGAPKEISE---TSV  1708

Qy    880  SKAITPKPQKPTKAPKKPTPSTIKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAML  939
              ||  :||                        |||:   |:  |  :|  ||
Db   1709  SKVLMSSP------PKKASSSKRASTLP------ATTLPSLKEASVLS----  1744

Qy    940  QTTTRPNQTPPNSKLVEVNPKSEDAGGAEGETP  971
                ||  :   |:|| |      ||  :  |||
Db   1745  -----PTATSSGKDSHISPVS-DACSTGTTTP  1770
```

RESULT 14
Q9VEL9

RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).

Query Match   11.1%; Score 756; DB 5; Length 2112;
Best Local Similarity   26.4%; Pred. No. 1.7e-42;
Matches 292; Conservative 107; Mismatches 367; Indels 338; Gaps 47;

```
Qy     21  QVSSQELSCKGRCFESFERGRECDCD------------AQCKKYDKC--------CPDY   59
              |:||:  ||: ||:| ||| |||                ||:||| |        |||
Db    953  QSSSSNQTSSSTTQKPFKPAEKCESEETFLADNENCSKFYRCVDNGKGGFTKVSFTCPPN 1012

Qy     60  ESFCAEVKD-NKKKNRFKKHKPPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTTSPKI  118
             :: |                  |  :|:|               |  ||:|    |:|| :|
Db   1013  TLWDPEANSCNHPDQIQTPKPLKCKKVVVSQGGSSSN------STSNSSSSSSNNSGSSSNSG 1066

Qy    119  TTAKPINPRPSLPPNSDTSKETSLTVNKETTV--ETKETTTTNKQTS----TDGKEKTT   171
             |:                  :|| :|        ::|::|||||||       |:|:||
Db   1067  SSSS----SSSSNSGSSSNTGSSSNSGASSSGGGSSNQGSSSNSGSSSGGSGSNSSGNQSTS 1120

Qy    172  SAKETQSI-----EKTSAKDLAPTSKVLAKPTPKAETTTGP--------             208
             |:|:|:          :|   |  || |  ||  ||:|||
Db   1121  SSTSSSSSSSNNNQGSSSSSSSTSSRKPNPSETCKVNGQFIGDRSDCAKFYRCVDN      1180

Qy    209  ------------ALTTPKEPTTPPTTPKEPASTTPKEPT                       233
                        |:|||: |::    ||:||| |||:|
Db   1181  DRGGENMVPFSCGPGTVWDAQMQACNHWAVKECGCIAPPTTSTPPTTSR-PTTASTRSPS 1239

Qy    234  PTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTT----------  282
             |||   ||||||||||||:|:||||||||||||||||||||| |||||
Db   1240  DQTISRPTIGP----PTTARPVTTARPTTSSPTTASSSQTTSPVTQAPNTDGKCRSEGFMA 1295

Qy    283  ----------------TKEP----APT--------TTKSAPTTPKEPAP            303
                             ||:|     |:|        ||:|| | ||  |
Db   1296  DPNNCSKFYRCVRNNKGGFTSIPFQCCGAGTVWDQDLQTCNHHNFNNCSTGTESTTPKPC- 1354

Qy    304  TTPKKEPAPTTPKEPAPTTTPKEPTTPKEPAPTTPKEPAPTTAPKKPAPTTPKE        363
               |  |||| |   |||:    ::| || ||||| |:  |  |    |||||||
Db   1355  -EPATNGTTATSTSSTTTP--PTTTTDLPPTSTTGLP-PTTTTELPPT------       1403

Qy    364  PAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPTK--EPSP  420
             |||| |  |||||   | |||      |||  |||| |  |||:| | |||     |:|
Db   1404  PPTTTTRLPPTTTTSLPPTTTTGLPPTTTTGAQPTTTTLSSETETSTVTTSPESTTQPPS 1463

Qy    421  TTTTKEPAP----------------------------TTPK-------EPSP         432
             |||:|||                               |||       |:|
Db   1464  TTTMKPLPAGTTECTGEGYMADPEDCRKYYRCINAGASYRKVNFTCPKGTGWNEEVQTCDY 1523

Qy    433  -EPAP---TTPKKPAPTTP----KEPAPTTPKE-PAPTTTKKPAPTTAPKEPAPTTPKETA  483
              :|||    ||   |||||     |||||| |   |||| ::|||| |:||||||||||
Db   1524  VENIPRCSKLPAEPITTTTPSEESKDPGSTTPQSTDEPTTVTKPI-TKPTE-EPSTEKPQK 1581

Qy    484  PTTPKKLTPTTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTPTPEEPAPTTPKAAAPNTP  543
             ||| :: |||:| || ||| |||||| |||  | ||:||||| ||||||
Db   1582  PTTQYPEKPTTEE--PEKPQKPT-TTEYPQKPTTEEPTEPEKPQKPT-TTEYPQKPTTTT  1637

Qy    544  KEPAPTTPKEPAPTT-PKEPAPTTPKETA---PTTPK--GTAPTTLKEPAPTTPKKPA     595
             :||| || || |||  |||| |||       |||     | ||| |||||||
Db   1638  EEPTPEKPQKPTTTEYPQKPTTEEPTTTSIPGYNPTTPGYNPTTPIPVETTTSTPG     1697

Qy    596  PKELAPTTTKEPTSTSDKPAPTTPKGTAPTTPKEPAPTTPKE-PAPTTPKGTAPTTLKE   654
             |||| |||:|||:||:   ||||||| ||||||||||||||   ||||||| ||||
Db   1698  YK---PTTTGEPTTTT-LPSTTTDAIQEPTTSKKPEPTTTTESPESSTPEGSVTTLQPE   1753
```

```
Qy    655 PAPTTPKK-------------PAPKELA------PTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTT 703
              ||||                ||:||                :|:|            |:|:||||
Db   1754 POPNYNCSSEGFFPDPEDCSRYYRCVDAAKNGKYQVYAFRCGKGTGTVWDISTETCNYADQV 1813

Qy    704 TPKKPAPTTPETPTPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPIPKALENSPKEPG 763
          ::||:|
Db   1814 SGN----------CSSGQTTTPGTTTEPGTTTESTTSSGKP----ETTSKAPEN----- 1852

Qy    764 VPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTTESKITATT 823
                   |  || ||                  :| |||           ||
Db   1853 ---TTTWA--PETTTT----SSPETTTT--VASETTTTTSGT--TT 1885

Qy    824 TQVTSTTTQODTTPFKITTLKTTTLLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKAT 883
              :|||:: |    |||      :  :|| | |||||||||||||||||||||||||||
Db   1886 TATPETTTTKPPKP-------ETTTIAGEETSTSKSPTTTE------------ 1918

Qy    884 TPKPQKPTKAPKKPTSTKKPKTMP 907
             |:|                |:|
Db   1919 SPAPSTNTSAP------CPETGP 1935


RESULT  15

ID   Q9XDH2     PRELIMINARY;   PRT;   763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH MUCIN HOMOLOG.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   Espitia C., Laclette J.L., Mondragon-Palomino M., Martens A.,
RA   Zhang X., Moreno C., Singh M.;
RT   "Cloning and characterization of a new member of the PGRS family that
RT   is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR003882; Pistil_extensin.
     PRINTS; PR01222; ATROPHIN.
     PRINTS; PR01217; PRICHEXTENSN.
     PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE  763 AA;  39168EC45A5916F8 CRC64;


Query Match       10.5%;  Score 715;  DB 2;  Length 763;
Best Local Similarity  28.2%;  Pred. No. 3.4e-40;
Matches 248;  Conservative 39;  Mismatches 361;  Indels 232;  Gaps 39;

Qy    197 PTPKRAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTTPKEPAPTTIKSAP 256
          | | ||             ||||                    ||
Db      3 PVP-------------APRALAPLPPAPPAPAEPKSKPPPPAP----- 46

Qy    257 TTPKEPAPTTTKEPAPTTPK----EPAPTTKSAPTTTKEPAPTTPKKPAPTTPKE 316
            |                 |:|          ||
Db     47 PCP--PAP------PAPPKPKSKAPFPPVPAPPAREIAPPLP--PAP----PEAPRE 90

Qy    317 PAPTTTPKEPTPTT----PKEPAPTTKEPAPTTPK-EPAPTAPK-KPAPTTPKEPAPTTPK 370
             | :|         |||  |:            |  ||:|      ::||
Db     91 SRPALPPCPPPPVVIPDPEPAPAVPPAPNSPPPPFPPPPPAPKFVPAPPVP--PVNSPP 148

Qy    371 EP--APTTTKEPSPTTPKEPAPTTKSAPTTTKEPAPTTPKEPSPTTTKEPAP 428
          :|   |      ||| |         ::|        ||       |||:
Db    149 FPPPPAALNPPAP-----PAPPLANSPPLPPAPPTPAGT--PPAAPWPPVPAAPKSKPA 201

Qy    429 TTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAP----TTTKKPAPTTPKEPAPTTPKETAP 484
```

```
Db    202 SPPRPPAP-----  ||| PMPATPMEFPPLPPVPPDPISKETPPAPPAPPIPAPVPIPPVPPLP 256
                     :|: |||         |    ||  ||     ||  ||   |||  ||  ||   |
Qy    485 TTPKKLTPTTPEKLAP------TTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTT-- 534
          ||:|:|             ||:|   |  ||  |:|  ||  ||   ||   ||  |:
Db    257 PVPNKIPPAPP--APPVAVAAVLVAPCPPLPPLPRNNHPPAPPAAPVPGVPLAPLPNSHP 313
Qy    535 ---PKAAAPNTPKEPAP-------------------------------- 548
             |:|
Db    314 PAPPSAPVPGVPLAPLPISGRPVSVWRGSFTTTLSTFCCRVCSGGEVLAGALNPSRPSRSPL 373
Qy    549 --TTPKEPAPTTPKEPAPTTPKETA---PTTPKGTA---PTTLKEPAPTTP-KKPAPKE 598
            ||||   |:||              |||:|       ||||     |||  ::|||:
Db    374 TTTTPALPAPIPPLPLPLPLPLPLPINTAVPPIPPLPPVTALAPLPPLPAPLPISPGVPPAP-- 431
Qy    599 LAPTTTKEPTSTTTSDKPAPTTPKGTAPTTTPKEPA----PTTPKEPAPTTPKGTAPTTL 652
             |||   ||:|            | |||:|||      |||         | |:||||
Db    432 -PIPGCKPWTTPLPAPAPPEPK-TVPVLPPGCPPSEKPNPPAPPEPPEPKSSPALPP 488
Qy    653 KEPAPTTP---KKPAPKELAPTTTTKGP-TSTTSDKPAPTTPKET---APTTPKEPAPTTP 705
          :||||       ::|||:|      |||  |:|        |||      ||||   |||
Db    489 APPAPSMPSAVRVPPSPPIPPAPPAARPASMPALPPAPPSPPATRLCPPLPPSPPAPNSP 548
Qy    706 KKPAPTTPETPTPPPTTSEVSTPTTTKEPTTIHKSP----DESTPELSAEPIPKA--LENSP 759
          ::||||   | | || |  :|      ||   ::|        |:|   |  ||   |:|||
Db    549 -PAPPAPPTPKLLS--ANPPCPVPPAPNRPPAPPAPPAPPELPAPPDPPTPPVANSP 604
Qy    760 KEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTTESKI 819
          :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    605 --PAPPAPPAPPSALPFVNPPA-----------------PPTPAAPK----------SRPAL 637
Qy    820 TATTTQVTSTTTQODTTPFKITTLKTTTLLAPKVTTTKKTITTTEIMNKPEETAKPKDRATN 879
            |||:                    |||      ||  |:           |||
Db    638 PAAPPAPPAPPVRATTP------------------------------PPAPPAPPAPN 665
Qy    880 SKATTPKPQKP-----TKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTM-----PELNP 930
             |||||  ||       |||                               |||||
Db    666 SMALPPAPPDPPIPLLIATPPAPPAPPLPMSPPAPPLPPAAPDPPAPPLTINQPPSPPLAP 725
Qy    931 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGET 970
           |:            |:  |:
Db    726 VPGAPLAPLPINGRPVFARKNSLI--------GSSSGDT 756
```

Search completed: April 26, 2002, 16:29:32
Job time: 587 sec

aa3.rspt

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:20:05 ; Search time 114.61 Seconds
(without alignments)
816.844 Million cell updates/sec

```
Title:          AA1
Perfect score:  6568
Sequence:       1 MAWKTLPIYLLLLLSVFVIQ..........ARAITTRSGQTLSKVWVNCP 1229

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters:   219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      PIR_68:*
         1:     pir1:*
         2:     pir2:*
         3:     pir3:*
         4:     pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1062.5	16.2	3020	2	A43932	mucin 2 precursor,
2	950	14.5	1664	2	T18262	S-layer protein -
3	853	13.0	1489	2	T31108	cyst germination s
4	802	13.2	1274	2	T16251	hypothetical prote
5	789.5	12.0	1367	1	S48478	glucan 1,4-alpha-g
6	784	11.9	1188	2	T30826	extensin-like prot
7	778.5	11.9	2187	1	S49915	nascent polypeptid
8	672	10.2	1344	1	A35175	mucin 1 precursor,
9	671	10.2	1151	2	T18535	high molecular mas
10	659.5	10.0	1229	2	T25697	hypothetical prote
11	635	9.7	7962	2	I38346	elastic titin - hu
12	633	9.6	489	2	T11622	extensin class 1 p
13	632	9.6	761	2	C84672	hypothetical prote
14	629	9.6	3507	2	T34513	hypothetical prote
15	627	9.6	990	2	I51618	nucleolar phosphop
16	625	9.5	6642	2	T29757	protein UNC-89 - C
17	622.5	9.5	839	2	F75518	hypothetical prote
18	622	9.5	3570	2	T45025	mucin MUC5B, trach
19	620.5	9.4	971	2	T19431	hypothetical prote
20	607.5	9.2	801	2	T29018	hypothetical prote
21	607.5	9.2	924	2	S27923	gene LF3 protein -
22	605	9.2	379	2	S50125	larval glue protei
23	562	8.6	1630	2	A53577	ascites sialoglyco
24	559.5	8.5	350	2	S22456	hydroxyproline-ric
25	556.5	8.5	856	2	T16543	hypothetical prote
26	556.5	8.4	875	2	S23760	polyphenolic adhes
27	553	8.4	2232	2	T34434	hypothetical prote
28	551	8.4	620	2	S06733	hydroxyproline-ric
29	545	8.3	873	2	A47283	calphotin - fruit
30	542	8.3	369	2	S20500	hydroxyproline-ric .
31	538.5	8.2	416	2	JU0465	extensin precursor
32	533	8.1	1087	1	QFMSH	neurofilament trip
33	532	8.1	756	2	T27642	hypothetical prote
34	530.5	8.1	1162	2	JH0557	hypothetical prote
35	530	8.1	865	2	A47282	exo-alpha-sialidas
36	522	7.9	328	2	JQ0985	calcium-binding pr
37	522	7.9	1459	2	T32271	hydroxyproline-ric
38	518	7.9	813	2	S70795	hypothetical prote
39	518	7.9	866	2	T45462	membrane glycoprot
40	518	7.9	1072	1	A37221	neurofilament trip
41	512.5	7.8	867	2	T45463	vsaA protein precu
42	509.5	7.8	867	2	A45155	mucin FIM-C.1 - Af
43	508	7.7	1832	2	T31113	mucin-like glycopr
44	506.5	7.7	700	2	A54641	interspersed repea
45	504.5	7.7	606	2	A43427	neurofilament trip

ALIGNMENTS

RESULT   1

```
A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
N;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C;Accession: A49963; A45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A;Reference number: A49963; MUID:94132002
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A;Reference number: A45106; MUID:93016075
A;Accession: A45106
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A;Status: not compared with conceptual translation
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A;Experimental source: colon
A;Molecule type: mRNA
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A;Reference number: A43932; MUID:91358717
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350,'L';1352-1411,'S';1413-1448,'P';1450-1503,'T';1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIN:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A;Reference number: A33532; MUID:89197956
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
```

```
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Experimental source: bronchus
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A;Reference number: PQ0328; MUID:92198477
A;Accession: PQ0328
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GDB:120203; OMIM:158370
A;Gene: GDB:MUC2
A;Cross-references: GB:M86523
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C;Keywords: glycoprotein; intestine; tandem repeat
F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>
```

```
Query Match               16.2%;      Score 1062.5; DB 2;    Length 3020;
Best Local Similarity     29.2%;      Pred. No. 8.3e-40;
Matches    342;  Conservative  74;  Mismatches  442;  Indels  315;  Gaps   39;

Qy    56  FKVTTPDFSTTQH---NKVSTSPKITTAKPINPRPSLPPNSDDTSKETSLTVNKETTVETK  112
              ::|       :   :|||     |:   ||||||       |  ||||||||  ||
Db  1240  WEICGPNGTVEKHFNICSITTRPSTLTFFTTIILPTTP--------TSFTTTTTTTTTPTS  1291

Qy   113  ET--TTTNK-----------------QTSTDGKEKTTS--AKETQSIEKTSAKD-------  145
          :|   |||                 :    |||     |||  |||
Db  1292  STVLSTTPKLCCLWSDWINEDHPFSSGSDDGDREPFDGVCGAPEDIECRSVKDPHLSLEQH  1351

Qy   146  -------------------------LAPTISKVLAKPTPKAE  161
                                   | |:  | || ||
Db  1352  GQKVQCDVSVGFICKNEDQFGNGPFGLCYDYKIRVNCCWPMDKCITTPSPPTTTPSPPPT  1411

Qy   162  TTTKGPALTTPKEPTPPTTPKEPAPTTTKPKEPAPTTIKSAP--TTPKEPAPTTTKSAPTT  220
          ||| | || ||||||||||||| | ||||  |||||| |||  ||||||||||||
Db  1412  TTTTLPPTTTPSPPTTTTTTTTPPPTTTPSPPITTTTTTTPSPPISTTTTTPPPTTTPS  1471

Qy   221  PAPTTTKEPAPTTTKEPAPTTTKSAP--TTPKEPAPTTPKKPAPTTPKEPAPT  279
          ||||||||||| | |||||||| |||  ||||||| | ||||||||||||
Db  1472  P-PTTTPSPPTTTTTTPSPPT--TTTTTTPPPTTTPPASTTTLPPTTTTPSPTTT  1529

Qy   280  TPKEPTPTTPKEPAPTTKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTIKEP  339
          ||||||| ||||||||| || |   || ||||||||||||||
Db  1530  TTTPPPTTTPSPPTTTPTTP-PTSTTTLPPTTTPSPPTTTTPPTTTTPPTTTTPSP  1588

Qy   340  SPTTPKEPAPTTTKSAP--TTTKEPAPTTTKSAP--TTPKEPSPTTTKEPAPTTTKEPAPTT  397
          |||||||||||| |||  |||||||||| |||  |||||||||||||||||
Db  1589  PTITTTPPPTTTPSPPTTTTTTTTPPPTTTPSPPTTTTTLPPTTTPSPPPTTT  1648

Qy   398  PKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTTAPKEPAPTTPKETAPTTPKKLPTTTPEKL  457
          ||| ||||||||||| |||||||  ||  |||||||| |   |||||
Db  1649  TTPPPTTTPSPPTTTTPSPPITTTTTTPPPTTTPSPPTTTTPSSPITTTTT-SPPTT  1700

Qy   458  APTTPEKPAPTTPEELAPTTPEEPTTPEPTTPKEAAAPNTNPKEPAPTTPKEPAPTT  517
          |||    ||||||||||||| |||||| |  ||||||||||||||||||
Db  1701  PTTTPSSPITTTTTTPSSTTTP-SPPPTTMTTTSPPTTTMTTTLPPTTTTSSPLTTT  1759

Qy   518  PKEPAPTTP---------KETAPTTP----  536
          ||||||||         |||||||
Db  1760  PLPPSITPPTTSPFSTTTPTTPCVPLCNWTGWLDSGRKPNFHKPGGDTELIGDVCGPGWAA  1819
```

```
Qy   537  ------------------------------------------------------------  536
Db  1820  NISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPMAFCLNYEINVQCCECV  1879

Qy   537  TAPTTLKEPAPTTPKKPAPKELAPTT--TKEPTSTTSDKP-----APT-TPKGTAPTTPKE  589
          ||||     |||  |      |||||   |||| || |    ||| ||| |
Db  1880  TQPTTM---TTTTTENPTPTPITTTTTVTPTPPTTTPSTQSPNGLQAPTTPISTTTTVTPT  1936

Qy   590  PAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTST-TSDKPAPTTP  648
          |||||||||||||   |||||| |||||||  |      |||| || |   ||||||
Db  1937  PTPTGTGTPT-TTPITITTTVTTTPTPTGTQTPTTVLITTTTTTMTPTTPTSTKSTTVTP  1995

Qy   649  KETAPTTPKEPAPTTTPKKP------APTTPETTPPPTTSEVSTPTTTTK-----EPTTIHK  696
          ||||||||||||||||||      |||||||   |||||||||        ||||||
Db  1996  ITTTTTVTATPPTPTGTQTPTPIMIPISTTTTTVTPTPPTTTPGTGTGPPPTHTSTAPIAE  2055

Qy   697  SPDESTPELSAEPTPKALENSPKEPCGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETT  756
          |:     |   ||| |||     |||    ||   ||  |  |     ||||    |||
Db  2056  PPESSTPQTSRSTSSPLTESTTTLLTSTLPPAIEMTSTAPPSTPTTAPPTTTSGGHTLSPPPST  2115

Qy   757  TAAPRKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITT----LKTTTLAPKVTTT  813
          ||   ||||   || |  |:|   |: |  |||:|||||||||     |||| ||| ||
Db  2116  TTSPPGTPTRGTTTT--GSSSAAPTTPSTVQTTTTSAWTTPTPPLSTTPPSIIRTTGLRYPSSV  2173

Qy   814  KKTITTTTELMNKPEE----------------------------TAKPKDRATTNSKATTP  844
          ||||||||        |                          |   |:|||||||||
Db  2174  LICCVLNDTYYAPGEEVYNGTYGTDTCYFVNCSLSCTLEFYNWSCPSTPSPPTTPSK-STP  2232

Qy   845  KPQKPTKAPKKPTSTHKKPKIMPRVRKPKTTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP  904
          |||||||||||||||||||| ||||||||||| ||||   || || ||||||||||||
Db  2233  TPSKPSSTPSKPTPGTTKPPECPDFDPPR----------QENETWWLCDCFM-ATCKY  2278

Qy   905  NQTPNSKLVEVNP-----------------------------KSEDAGG----------AEGETPHMLL------  934
          ||||||||                                 |||         | |||||
Db  2279  NNTVEIVKVECEPPPMPTCSNGLQPVRVEDPDGCCWHWECDCYCTGWGD--PHYVTFDGLY  2337

Qy   935  -----RPHVFMPEVTPDMDYLPRVPNQGIIIN  961
               | |:|   |  |:|:   |  |: |:
Db  2338  YSYQGNCTYLVEEISPSVD------NFGVYID  2364
```

```
Query Match               14.5%;      Score 950;  DB 2;    Length 1664;
Best Local Similarity     31.0%;      Pred. No. 4.4e-35;
Matches    312;  Conservative  104;  Mismatches  361;  Indels  230;  Gaps   55;

Qy   152  VLAKPTP-KAETTTKGPALTTPKEPTPPTTPKEPAPTTTTIKSAPTTPKEPAPTT  210
          |: ||:|  |:|| ||||||||||||   |||||||| |||
Db   758  VVIQPAPIKAASDEPIPTDTPSDEPTPS----------DEPTPS---DEPTPSDEPTPSD  804

Qy   211  TKSAPTTPKEPAPTTTKEPAPTTPKEPA---PTTTKEPAPTTTKSAPTTPKEPAPTTPK  266
          |:||||||||||||| |||||||||||   |||||||||| |:||:||||||||||
Db   805  EPTPSETEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTP  864

Qy   267  KPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTAPAPKKPAPTTPKEPAP  325
          ||||| ||  ||||  :|| ||||||||||||||  ||||||||||||
```

```
RESULT    3
T31108
cyst germination specific acidic repeat protein precursor - Phytophthora infestans
C;Species: Phytophthora infestans (potato late blight agent)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31108
R;Reference number: 220986
submitted to the EMBL Data Library, April 1998
R;Goenhardt, B.
C;Accession: T31108
A;Accession: T31108
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1489 <QOE>
A;Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC72308.1

C;Genetics:
A;Gene: car90

Query Match    13.0%;   Score 853;   DB 2;   Length 1489;
Best Local Similarity   32.3%;   Pred. No. 7.9e-31;
Matches  355;  Conservative  46;  Mismatches  475;  Indels  222;  Gaps  49;
```

```
Qy    33 TKKKPKTPKPPVVDEAGSGLDNGDRFKVTIPDTSTT-----QHNKVSTSPKIITPINPR   86
Db   338 TPYAPTEKPYDVEETTYVTEESTYAPTKSETNAPTERMHYAHIEKPCDFEVTMYAPTEET  397

Qy    87 PSLP-------PNSDT-----SKETSLIVNKEITVEFKETTT----TNKQTSTD        124
Db   398 TYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEETTYAPTEETTYAP  457

Qy   125 GKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPA  184
Db   458 TEETTYAPTEETPYEPTEETTYAPTEETTYAPT---EETTYASTEETTYAPTEETTYAPA  514

Qy   185 STTPKEPTPTTKSAPTTPKEPA---TTKSAPTTPKEPAPT--TTKEPAPTTPKEPAPTT   241
Db   515 EETTYEPTEET--TYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETT        573

Qy   242 TKKEPAPTTPKSAPT------TPKEPAPTTPKKPAPTTYAPTEETTYAPTTPKE        291
Db   574 TYAPTEETT-YAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYA  632

Qy   292 P-------AP-------TTKEPAPTTPKEP-----APTAPRKPAPTTPKEPAPTT        328
Db   633 PTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEE         692

Qy   329 KEPAPT--TTKEPSPTTPKEP-----APT-TTKSAPT-----A                   363
Db   693 TTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYA                 752

Qy   364 PT-TTKSAPT--TPKEPSPTTTKEP------APTTPKEPAPTTPKKPAPTTPKEPAPT--  412
Db   753 PTEATTYAPTEETTYAPTEETPYEPTEETTYAPTEETTYAP                    812

Qy   413 TPKEPAPTTPKKPAPTAPKEPAPTTPKEPAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEE  472
Db   813 TPYEPTEETTYAPTEETTYAPAEETPYEPTEET---TYTPTEETTYAPTEETTYAPTEETT 870

Qy   473 LAPT--TPEEPTPTPEEP-----APT-----TPKAAAPNTPKEPAPTTPKEP-         513
Db   871 YAPTEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYASTEETTYAPAEETPYEPT  930

Qy   514 -----APTTPKEPAPTTPKETAPTTPKGTAPT---TLKEPAPTTPKRKPAPK-ELAPT--TT 563
Db   931 EETTYAPTEETTYAPTEETMYAPIEETTYAPIEETTYAPTEETTYAPTEETTYAPTEETT  990

Qy   564 KEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP-----APTTPKGTAPT-----     609
Db   991 YAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAST 1050

Qy   610 -----TLKEPAPTTPKKP------APKE---LAPT--TTKGPTSTTSDKPAPTTP       648
Db  1051 EETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETT 1110

Qy   649 KETAPTTPKEPAPTTPTPTPPPTTPSEVSTPTTTKEPT--TTIHKSPDESTPELS      706
Db  1111 YAPAEETPYEPTEETTYAPTEETMYAPIEETTYAP-APTEATTYAP              1169

Qy   707 AEPTTPKA-LENSPKEPG----VPTTKTPAATKPEMTTTAKDKTTERDLRTTP---ETT   756
Db  1170 TEETPYAPTEETTYEPTGTTYAPTEETTYAPTEETTYAPTEETPYEPTEET         1229

Qy   757 TAAPKMTKETATTTEKTTESKITATTTQVSTTTQQDTT-----PFKITT---LKT      803
Db  1230 TYAP-----TEETTYEPTEETTYAPTEETTYAPTEETTYAPIDETTYGPTEE        1284

Qy   804 TTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAT--NSKATTPKP-QKPTKAPKKPT---  857
```

```
Db   865 SDEPTPSDEPTPS--DEPTPS-DEPTP-SDEPTPSSETPEEPIPTDTPSDEPTP        919
Qy   326 TTPKEPA--PTTTKEPSPT-TPKEPAPTTKS-APTTTKEPAPT---TTKSAPTTKEPS    378

Db   920 SDEPTPSDEPTPSSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPT            979
Qy   379 --PTTTKEPAPT-TPKEPAPTTPKKPAPTTDTPSDEPTTKKPAPTAPKEP           433

Db   980 PSDEPTPSDEPTPSSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTP         1039
Qy   434 APT---TPKETAPTTPKKLTPTTPEKLAPTTPEKPAPT---TPEELAPTTPEEPTPT-TP  486

Db  1040 TPSDEPTPSE--TPEEPIPTDTPSDEPTPSSDEPTPSD-EPTPSDEPTPSDEPTP      1094
Qy   487 EEPAPTTPKKAAAPNTKEPAPTTPKEPA--PTTPKEPAPT-TPKEPAPTTPKKPAPTTKEP 543

Db  1095 EEPIPTDTPSDEPTPSSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSD          1154
Qy   544 EPAPTTPKKPAPKELAPTTTKEPT-STTSDKPAPTTPSDEPTPSETPEEPTPSETPEEPTI 602

Db  1155 EPTPS--DEPTPSD-EPTPSDEPTPSSETPEEPIPTDTPSDEPTPS--DEPTPS-     1208
Qy   603 PKGTAPTTLKEPAPT--TPKKPAPKELAPTTTKGPTSTTSDKPAPT---TPKETAPTTPKE 658

Db  1209 ---DEPTPSDEPTPSETPEEPI-                                      1253
Qy   659 PAPTTPKKPAPT-TPETPPPTTSEVSTPTTIHKSPDESTPELSAEPTTKALENS       717

Db  1254 PTPS--DEPTPSETPEEPIPTDTPSDEPTPSD----PSDEPTP--SDEPTPSDEPTP    1305
Qy   718 PKEPGVPTTKTPAAT---KPEMTTTAKDKTTERDLRTTPETTPAAPKMTKETATTTEKTTE 775

Db  1306 SDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSD----EPTPSDEPTPSDEPTPS      1361
Qy   776 SKITATTTQVSTTTQQTTTPFKITTLKTTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDR 835

Db  1362 DEPTPSDEPTPSETPEEPTP----TTTPTPTSTT-                         1396
Qy   836 ATNSKAP-----TPKPQRPTKAP--KKPTSTKKRKTMPRVRKPKTTTPTPRKM-----   880

Db  1397 SGGSGGSGSGGGGGGGGGGGGTVPTSPTPTPTSKPTSTPAP---TEIEEPTPSDVPGAIGGEHRA 1453
Qy   881 --TSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHM       932

Db  1454 YLRGYPDGSFRPERNITRAEAAVIF----AKLL----GADESYGAQSASPYSD       1498
Qy   933 LLRPH-----VFMPE-----VTPDMDYLPRVPNQGI-----                    958

Db  1499 LADTHWAAWAIKFATSQGLFKGYPDGTFKPDQNITRAEFATVLHFLTKVKGQEIMSKLA  1558
Qy   959 ---IINPMLSDETNICNG-----KPVDGLITL-----RNGILVAFRGHYFWMLSPFSPPS 1005

Db  1559 TIDISNPKFDD-----CVGHWAQEFIEKLTSLGYISGYPDGT                  1600
Qy  1006 PARRITEVWGIPSPIDTVFTRCNCEGKTFFFKD--SQYWRFTNDIKD            1050

Db  1601 YIKRSESV---ALINRALERGPLNGAAPKLFPDVNESYWAF-GDIMD            1642
```

```
Db  1285  TTYAPTEATTYAPTEETTYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPMEETPYE 1344
                  ||| :| | |:  |::| |  |  | |||  :
Qy   858  ------------STKKP-KTMPRVRPKTTPPRKMTSTMPELNPTSRIAAAMLQTTTRP-N 905
Db  1345  PAEESTSTSVSTEKRPCNTEEFTDEPTDEPTDE--PSDEPTDEPTDEPTDLPTDEPSTPCDN 1402
Qy   906  QTPNSKLVEVNPKSEDAG 923
Db  1403  QGINGIGVENKVRYNNAG 1420

RESULT  4

T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C;Accession: T16251
R;Leimbach, D.
A;submitted to the EMBL Data Library, January 1996
A;Description: The sequence of C. elegans cosmid F35A5.
R;Reference number: Z18485
A;Accession: T16251
A;Molecule type: DNA
A;Residues: 1-1274 <LEI>
A;Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GN00
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Experimental source: strain Bristol N2; clone F35A5
C;Genetics:
A;Gene: CESP:F35A5.1
A;Map position: X
A;Introns: 1272/2

Query Match    12.2%;  Score 802;  DB 2;  Length 1274;
Best Local Similarity  28.0%;  Pred. No. 1.2e-28;
Matches  290;  Conservative 102;  Mismatches 414;  Indels 230;  Gaps 56;

Qy   30  KNRTKK--KP------TP------KEPVVDEAGSGLDNGD----FKVTTPDTSTTQHNKVSTS 74
Db  274  KNPTKKWKPPWEDETPVEVKEPVPEKKAPVLKKKDPAPAAKARDPSEPSKAAPKKVEPS 333
Qy   75  PKITTAKPI-NPRSLPENSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAK 133
Db  334  SPVVPPTPVKNPVKKIKPPWEVDDEPAEEVKKPSAPEKKTPVLKRKEEPEPSSTTPSSSDPS 393
Qy  134  ETQSIEKTSAKDLAPTSKVLAKPTPKAE-----TTTKGPA----------------LTTP 172
Db  394  PKKAAPAVKPRDSSPKKATPLQADPKAQEVPPTPVKNVKYKPPWEVDDEDPVEEVKQP 453
Qy  173  KEPTPTTP-----KEPASTTPKEP----TP-TTIKSAPTTPKEPAPTTTKSAPTTPKEPAP 223
Db  454  EAPAKKTPVLKRKEPAAKDTAKPATSKIPETPEKKDPVKPRDSSPKKVAAKPDSAQAPA- 512
Qy  224  TTTKEPA-------PTTPKEPAPTTPKEPAPT------TTKSAPTTP 257
Db  513  TPVKNPVKKWRPPWEDDETPADDVSKPTDAKKTPSLAKKDPAPAKESLKPKADTKAPAKP 572
Qy  258  KEP-------APTTPKKRPAPTTPKE-----SPTTPKEPAPTTPKEP 307
Db  573  RDPSPKKVAPTAPEKKTPVLAKKEPAGPADSKIKEPEKSKPRDSPSKKVPAKPVKTEV 632
Qy  308  APTTAPKKPAP------TTPKEPAPTTPKEPAPTTTKSAPTT 357
Db  633  APAAVKPEPISKPKDTAPKKAEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPE 691
Qy  358  TTKEPAPTTTKSAPTTP-KEPSPTTTKEPAPTTPK--EPAPTTPKKPAP--TTPKEPAPT 412
Db  692  PEKK-TPVLAKKAPTKPDSEAAADPVSGGPSSKDPKLAKKAPAVKPRDPSPMKAVPIKPAPK 750
Qy  413  TPKEPAPTTTKKPAPTA-PKEPAPTTPKETAPTTPKKLTPTTP----EKLAPTTPEKPA 466
Db  751  T--EVPPAVVKKPEPVAKSRDPSPKKAK-AEPNSP--VVPPTPVKNPVKKWKPPWEDDDA 805

Qy  467  PTTPEELAPTTPEEP-------TPTTPEEPAPTTPKAAAPNIPKEPAPTTPKEPAPTTP-PKE 520
Db  806  PAEPVNPEPEKKTPVLAKKTPVKPRDPSPKKAVPAKPSTKTDAPPVSVKKPEPVSVKPKE 865
Qy  521  PAPTTPKETAPTTPKGTAPTTPLKEPAPTTPKKPAPKELAPT-TTKEPTSTTSDKPAPTTP 579
Db  866  PSPKKAEPNSPVVP----PTPVKNPVKKW-KPPWEDDDDEPTEEVKKPSE--PEKKTPVLA 918
Qy  580  KGTAPTTPKEPAPTTPKGTAPTTPLKEPAPT------TPKKPAP------KEL 627
Db  919  K-KEPEKPKD-APKVAAKPRDPSPKKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEA 974
Qy  628  APTTTKGPTS------TTSDKP----APTTPKET------APTTPKEPAPTTPKKPAPT 670
Db  975  PPTPVKNPVKKWKPPWEDDDEPAEEVSAPEPEKKTPVLAKKAPAKPRDP--SPKKAAPV 1031
Qy  671  TPETPEPPTTSEVSTPTTTKEPTTIHKSP----DESTPELSA-EP---TPRALENSPKEPG 722
Db  1032 AAK-PDPKIPEV-PPTPVKNPVKKWKWKPPWEDDDEPSEPVSAPEPEKKTPVLAKKAPTKPA 1089
Qy  723  V-----------PTTKTPAAT-----KPEMTTTAKDKTTTERDLRTTPETT--TAAPK 761
Db  1090 TKPDSEAAADPVSGPTSKDPKLSKKAPAVEKRPKPTTDPKDKDKLKPSPAKKPEKAPEPAAPK 1149
Qy  762  MTKETATTTEKTTTESKIITATTTQVTTSTTTQDTTPFKITTLKTTTLLAPKVTTTKTITTTE 821
Db  1150 KWKPVWDDDDEPEEADFTVPAPSKKPDTEDPADPLG--------GPKTKDPK------- 1193
Qy  822  IMNKPEETAKPKDRATNSKATTPRPGQKPTRAPKKPTSTKKRPVKPKTMPRKVRKPK-- 872
Db  1194 -LNKKAPAEKPTEK--------PKPKEVSKEPPKPTEPPKP-AAPKKWKPPWEDDDEPE 1243
Qy  873  ---TTPTPRKMTSTMP 885
Db  1244 ADFTMPAPKKPDTEDP 1259

RESULT  5

S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR01
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
A;submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN000009;
A;Molecule type: DNA
A;Reference number: B26877
A;Accession: B26877
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Reference number: A91831; MUID:87194600
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacc
A;Reference number: S27281; MUID:89031230
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
```

Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A:Reference number: JC6123; MUID:96323237
A:Accession: JC6123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAM>
A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C:Genetics:
A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MIPS:YIR019c; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match   12.0%;   Score 789.5; DB 1;   Length 1367;
Best Local Similarity   28.4%;   Pred. No. 4.7e-28;
Matches 301; Conservative 104; Mismatches 483;   Indels 171; Gaps 47;

Qy   59 TTPDTSTTQHNKVSTSPKITTAKPINP--------RPSLPPNSDTSKETSLTVNKETTVE 110
          : ::||| || || | :||| | | | || ||||
Db 236 STSESSTTTSSTSSESSTSSSSTTAPATPTTTSCTKEKPTPPTTTSCTKEKPTPPHDTTPC 295

Qy 111 TKETTTTNKQTSTDGKEKTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETT--TKGPA 168
          ||: ||| :|||: |: ||||: :: | || : | || |: ||||
Db 296 TKKKTTTSK-TCT---KKTTTPVPTPS-SSTTESSSAPV------PTPSSSTTESSSAPV 344

Qy 169 LTTPKE------PTPTTPKEPASTTPKEPTPTTIKSAP----TTPKEPAPTTTKSAPTTP 218
           :|||| |||: || ||:| ||| | | |:| :|| || || || |: ||||
Db 345 TSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTPSSSTTE 404

Qy 219 KEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKEPAP 278
         ||| || |: || | || |: ||: || | | ||| | ||| ||| | || |||
Db 405 SSSAPVTS----TTESSSAPVTS---STTESSSAPVT---SSTTESSSAPVTSSTTES 453

Qy 279 TTPKEPTP---TTPKEPAPTTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAP---TTPKEP 331
          |||| | ||| | || ||:| |: | || | | || || || ||||
Db 454 SSAPVTPSSSTTESSSAPVT--SSTTESSSAP-VPTPSSSTTESSSAPVTSSTTESSSS 509

Qy 332 APTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAP----TTTKSAPTTPKEPSPTTTKEPAP 387
         || | || ||| || |: ||: | ||| | || ||| | ||| | ||| |
Db 510 APVPTPS-SSTTESSSAPATPSSSTTESSSAPVTSSTTESSSAPVTSSSTTESSSSSTP 568

Qy 388 ----TTPKEPAPT-TPKKPAPTTPKEPAPTTKKPAPTAPKEPAPTTPKETA 442
             || ||| || ||| || ||| ||| || | | || || ||||| |
Db 569 VTSSTTESSSAPVPTPSSSTTESSSAPVT----PSSSTTESSSAPATPSSSTTESSSA 624

Qy 443 PTTPKKLTPTTPKEKLAPTTPKEPAPTTPEELAPT-TPEEPTPTTPEEPAP-----TTPKA 496
          ||||| | | ||| |: |||| || || | | |||| | | | || |||||
Db 625 PVT-----SSTTESSSAP-VPTPSSSTTESSSAPVTSSSTTESSSAPVTPSSSTTESS 679

Qy 497 AAPNTPKEPAPTTPKEPAPTTTKEPAPT-TPKGTAPTTLKEPAPTTPKKPAP 555
         | || ||||| || ||| |: || | ||||| |: |||| | |: ||
Db 680 SAPVT-----SSTTESSSAPVT---SSTTESSSAPVTPSSSTTESSSAPVTPSSSTTE 731

Qy 556 KELAPTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAP---TTPKEPAPT-TPKGTAPTTL 611
         || |||| |: || | | || | || || ||||| || || || |||| |||:|
Db 732 SSSAPVTPS-SSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTPSSSTTESS 790

Qy 612 KEPAPTTPKKPAPKELAPTTKGPTSTTSDKPAPTTPKETAPTTPKEPAP------TTPKK 666
         || || || | || |||: | || | |: || | || |||| | |||||
Db 791 SAPVPTPSSSTTESSVAPVPTPSSSTTESSTPFSSSTESSSVVPTPSSSTTESS 850

Qy 667 PAPTTPETPPPTTSEVSTPT-----TTKEPTTIHKSPDESTPELSABPTPKALENSPKEP 721
          | |||| | | || || | || || : | | | || | ||| |
Db 851 SAPVSSSTTESSVAPVPTPSSSSSNITSSAPSSL---PFSSTTESFSTGT-TVTPSSSKYP 906

Qy 722 GVPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITAT 781
         | |||| | ||: || ||: |:| |: | || | ||| |
Db 907 GSQTETSVSSTTETTIVPTKTTTSVTTPSTTTITTTVCSTGTNSAGETTSGGCSPKIVTTT 966

```
F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
F:1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
F:1-19,29-1344/Domain: signal sequence #link PREB #status predicted <PREB>
F:1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice f
F:138-1017/Region: 20-residue repeats (GSTAPPAHGVTSAPDTRPAP)
F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F:1245-1272/Domain: transmembrane #status predicted <TRM>
F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predic
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match        10.2%;  Score 672;  DB 1;  Length 1344;
Best Local Similarity   27.8%;  Pred. No. 7.6e-23;
Matches 310;  Conservative  78;  Mismatches 472;  Indels 256;  Gaps 56;

Qy  101 LTVNKETTVETKETTTTNK--QTSTDGKEKTTSAKETQSIEKTSAKD-LAPTSKVLAKPT 157
            |||  ||  | |||| ::   | | |||::|| :| | : ::|| :: ||| |   ||:|||
Db   15 LTVLTATTAPKPATVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS 74

Qy  158 P-KAETTKGP--ALTTKEP----------------------------TPTTTPKEPASTTP---------KE 190
             | ||||||  | |||||                            | || ||||||||          ||
Db   75 PSGGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

Qy  191 PTPTT------IKSAPTTPKEPAPTTT------KSAPTTPKEPAPTTTKEP-------A 230
               | | ||       :|| :|| |||:||          |||:|| :| | : :            |
Db  135 PAPGSTAPPAHGVTSAPDT--RPAPGSTAPPAHGVTSA 190

Qy  231 PTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTP----------KKPAPTTPKEP-------- 276
                |  ||| :|| :||   ||| :|| | ||| |||            |||| |||||
Db  191 PDT--RPAPGSTAPPAHGVT-SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 247

Qy  277 --APTTPKEPTPTTP----------------------KEPAPTTKEP------APTTPKEPAPTTPK---- 313
             | |||||  | ||                     ||| ||||||       | |||||| | || |
Db  248 TSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 307

Qy  314 KPAPTTPKEPAPTTPK-----EPAPTTPKEPSPTTPKEPAPTT---TKSAPTTTKEPAPTT 366
             | || |||| | || |       | |||||| |||||| | |||      |||| :|||| | ||
Db  308 TSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 367

Qy  367 TKSAPTTKEPSPTTTKEP--------APTTPKEPAPTTPK----KPAPTTPKEPAPTTP--- 415
             | :|| :|| | ||||||         | ||||||| | |||       |||| ||||||| |
Db  368 T-SAPDT--RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 424

Qy  416 ---EPAPTTTKKPAPTTPKAPTAPKEPAPTTPKETAPTTPKKLLPTTPEKLAPTTPEKPAPTTPE-- 471
                 | | |||    || || || : ||| |||||  | | |||||:|||| | | |||||||  | | | | ||
Db  425 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 484

Qy  472 ---ELAPTTPEEPTPTTP--------------EEPAP--TTPKA----AAPNTPKEPAPTTPK----- 511
               | |||||| | || |              ||| |  |||| |       | |||||| | || |
Db  485 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 544

Qy  512 ---EPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTP----------KKPAPKEL 558
                 | | |||| | || ||| | | |||| :|||| | | | |||||            |||||| |
Db  545 HGVTSAPDTRPAPGSTAP----PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGST 600

Qy  559 APTTTKEPTSTTSDKPAP--TTPKGTAPTTPKEPAPTTPKG---TAPTTLK 612
             | |||:|| | ||:|||   || || |||||| | |||| |   | |||| |
Db  601 AP-PAHGVTSAPDTRPAPGSTAPPAHGVTS----APDTRPAPGSTAPPAHGVTSAPDTR 655

Qy  613 EPAPTTP-----------KKPAPKELAPTTKGPTSTTSDKPAP---TTPK----ETAPTTP 656
             | | |||           || |||| |||| | | ||:| |||    ||||     ||| |||
Db  656 APGSTAPPAHGVTSAPDTRPAPGSTAP-PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTR 714

Qy  657 KEPAPTTP-----------KKPAPTTPETPPPTTSEVSTPTTTKEPTT--------IHKSPD 699
             ||| |||           || || || | ||||| | |||| |||| |         | |||
Db  715 PAPGSTAPPAHGVTSAPDTRPAPGS--TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD 772

Qy  700 ESTPELSAEPTPKALENSPKEPGVPTTKTPAA----------TKPEMTTTAKDR---TTERD 748
             | | || | || |   | | || || |||| |           ||  |||| |    |||||
Db  773 TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD 832

Qy  749 LRTTPETT------TAAPKMTKETATTTEKTTESKITATTTQVTSTTQDTTP--FKIT 799
             | ||| |||       | : |   |   ::|
```

A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129; GB:
A:Experimental source: splice form B
A:Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl-en
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel
J. Biol. Chem. 265, 15286-15293, 1990
A:Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli
A:Reference number: A35886; MUID:90368715
A:Accession: A35886
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-992,1033-1344 <GEN>
A:Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence
A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870
R:Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A:Reference number: A35887; MUID:90368716
A:Accession: A35887
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599
A:Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated seque
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
A:Reference number: S10571; MUID:90276413
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
R:Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A:Reference number: S40293
A:Accession: S40293
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WR2
A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
R:Abe, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A:Reference number: A36735; MUID:90088473
A:Accession: A36735
A:Molecule type: mRNA
A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543
R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H
J. Biochem. 112, 609-615, 1992
A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A:Reference number: S51026; MUID:95080414
A:Accession: PX0066
A:Contents: annotation
A:Molecule type: mRNA
A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037;1038-1057 <MAS>
A:Cross-references: GDB:120705; OMIM:158340
R:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
A:Reference number: JX0235; MUID:95123189
A:Contents: annotation
C:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region an
C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
C:Comment: partial repeats. The repeat shown is defined by SmaI nuclease sites.
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C:Genetics:
C:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-1q23
A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
F:1-344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
F:1-62/Region: mucin 1 amino-terminal non-repetitive

```
Db   569  KPVPLMALTPQPVTAQMVTQLAATKPSPIVPRASPK-ALMTPPPPPGLPRALAAAKLLG  627
Qy   475  -PTTP--------EEPTPTTP------EEPAPTTP----------KAAAPNTPKE-----  504
Db   628  LPSSPVASAMHAKVTPRPLPASPVPMAASPASLGPDAARVALATNAASPGAKPEAAGGNG  687

Qy   505  ----PAPTTPKEPAPTTPKEPAPTTPKETAPT----------TPKGTAPT-------  540
Db   688  TLMAPMGAANTQMAPIGAAGAAQTAPMGAAHTHVSPMGAGGATQMSPTGAANTHMSPIGA  747

Qy   541  --------TLKEPAPTTPKKPAPKELAPTTTKEP------TSTTSDKPAPTTP  579
Db   748  GGATQMSPMGAAANTQMSPMGATTQMSPMGAAAATTQPSPMGAAAATQVTATSAGNTMQVSP  807

Qy   580  KG--TAPTTPKEPAPTTPKGTAPTTLKEP--APTTPKKPAPKELAPTTTKGP  635
Db   808  MGAATPPQTPSVGAATTP-QPSPM---GAATTLMSPMGAATTPQ---PSPMGAVTTQPP  859

Qy   636  ----TSTTSDKP-APTTPKET---APTTPKEP---APTTPKEP--APTTPETPP-------  676
Db   860  PMAATNTTQPPMAASTPQSTPMGAATTQSPMGATTTQSPPMGCASTPQAPPTVAGSPT  919

Qy   677  PTTSEVSTPTTTKEPTTIHKSPDESTPEL--SAEPTPKALENSPKEPGVPTTKTPAATKP  734
Db   920  PPPIPPSPTAQTSRQPMSKSPPPDPKAPSAAAQTSPAAHVANASPGV-TAVSPA---P  975

Qy   735  EMTTTAKDKTTERDLRTTPETTTAAPKM--KETATTTEKTTESKITATTTQVTSTTTQDT  793
Db   976  IGVTEASPSADGARLSPGPTAATDGPKASPAATADVTEAATD--VTAAATAVPA----EA  1029

Qy   794  TPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAP  853
Db   1030 AP--------TKAKRSSSSSSSSSSSSSSSSSSSSSSSSDSDSSSSSSSESNPASPA---  1077

Qy   854  KKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLV  913
Db   1078 --------PAVGDGQQQMTPGAAQSVPP-------VTEAAVQ----------------  1104

Qy   914  EVNPKSEDAGGAEGE  928
Db   1105 EAAAAAAAAGAERE  1119


RESULT  10
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T25697
R;Fulton, B.
submitted to the EMBL Data Library, August 1996
A;Description: The sequence of C. elegans cosmid F16F9.
A;Reference number: 220071
A;Accession: T25697
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1229 <FUL>
A;Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9.2
A;Experimental source: strain Bristol N2; clone F16F9
A;Genetics:
A;Gene: CESP:F16F9.2
A;Map position: X
A;Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match     10.0%;   Score 659.5;   DB 2;   Length 1229;
Best Local Similarity   29.0%;   Pred. No. 2.5e-22;
Matches 307;   Conservative 78;   Mismatches 339;   Indels 333;   Gaps 52;

Qy   57   KVTT-PDTSTTQ--HNKVSTSPKITTA-KPINPRPSLPPNSDT-------SKETSLTVN  104
Db   119  KVTTSTDASTTNAPTTGKDSTTPEIITGIVINSKESVTDMSTTRFSTTLSPTTELLTS  178
```

of chick

```
Db   833  TRPAPGSTAPPAHGVTSAP---DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT  888
Qy   800  TLKTTTLAPKVTTTKKTITTT--EIMNKPEETAKPKDRATNSKATTPKPQKPFKAPKKPT  857
Db   889  SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAP--GSTAPPAHG  946

Qy   858  STKKPKTMPRVRKPKTTPTPRKMTSTMPELNP----TSRIAEAMLQT-TTRP---NQTPN  909
Db   947  VTSAPDTRP--APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP  1003

Qy   910  SKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETN  969
Db   1004 AHGVTSAPDTRPAPGSTAPPAH-----GVTSAPDNRPALGSTA----PPVHNVTS  1049

Qy   970  ICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPS  1005
Db   1050 ASGSASGSASTLVHNGTSARATTTPASKSTPFSIPS  1085


RESULT  9
T18535
high molecular mass nuclear antigen - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18535
R;Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A;Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A;Reference number: Z18955; MUID: 9803440
A;Accession: T18535
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1151 <SHI>
A;Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PID:BAA24137.1

Query Match     10.2%;   Score 671;   DB 2;   Length 1151;
Best Local Similarity   25.9%;   Pred. No. 7.2e-23;
Matches 284;   Conservative 100;   Mismatches 419;   Indels 292;   Gaps 55;

Qy   38   TPKP-------PVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTS--PKITTAKPINPRPS  88
Db   113  TPPPSQGPAGTPPPSQGAAGAPKGDGTAQPSGTKSGADGKPAAQDVPKATTA-ATEARP-  170

Qy   89   LPPNSDTSKETSLTVNKETTVETKETTTNK--QTSTDGKEKTTSAKETQSIEKTSAKDL  146
Db   171  ------ASAASPTVPKATAEATAVTAASQSOAPKAATDAAAVTAA--SQSAPKATV-EV  219

Qy   147  APTSKVLAKPTPKAETTTKGPALTTPKEPTP-TTPKEPASTTPKEP----TPTTIKSAPT  201
Db   220  KPAAAAVAKEAKAVTAAAAAPKATAEAKPAPVTSPTIPCSSAEAKPLTAASPTASKA--T  277

Qy   202  TPKEPAPTT---------TKSAPTTPKEPAPTTTK----------  227
Db   278  AEAKPVPATASLMATKVTAEAKPAPSPSVPKATTDIKAVITATAPKAGPDVKPAVAVCAEA  337

Qy   228  EPAPTTPKEPAPTTKEPAPTTTKSAPTT-------PKEPA----PTTPKKPAPTTPK  274
Db   338  KPAPPPPPQQLPKAAAAAPTGELKPATAPPHGSPRANSHTVTVTPPNVPRAAAATVP-  396

Qy   275  EPAPTTPKEPTPTT----PKEPAPTTKEPAPTAPKKPAPTTPKEPAPTTPKE  330
Db   397  -TAGAVPKASTGTPAAPAQQPVP---KAAPVTPPSPQQAVPRAATAAA----APVTPQQ  448

Qy   331  P---APTTTKEPSPTTPKEPAPTTPKEPAPTTKSAPTTTKEPAP---TTTKSAPTTPKEPSPTTTKE  384
Db   449  PVTKAATTNAPPPQPIPKAATTTTATPVTPQQPIPKAGTDAAPPAVPKAPSDGRAAT  508

Qy   385  P----APTTPKEPAPTTPKEPAPTTPKEPAP---------TTPKEPAPT-TTKPAPTAP  430
Db   509  PGVPNAATDPQKPPPTPQSVPSAVTEPKPQRAAPPSNEATPAVPSPSPNLKSPLPTIP  568

Qy   431  K-------EPAPTTPK--ETAPTTPKKLTPTTPEKLAPTTPEELA-------  474
```

```
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: I38346
R;Labeit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A;Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A;Reference number: A57430; MUID:96026330
A;Accession: I38346
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-7962 <RES>
A;Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:127867; OMIM:188840
A;Map position: 2q31-2q31

Query Match          9.7%;   Score 635;  DB 2;   Length 7962;
Best Local Similarity  22.8%;  Pred. No. 2e-20;
Matches 300;  Conservative 106;  Mismatches 463;  Indels 444;  Gaps 61;

Qy    18 VIQQVSSVKDNKKNR----------TKKKPTP----KPPV--VDEAGSGLDNGDFKVT----   59
Db  5814 VIQVQKEYEESHERKVPAKVPEKVIKKPVIEKIEKTSRRMEEEKVQVTKVPE           5873

Qy    60 ----------TPDTSTTQHNKVSTS-PKITTAKPINPRSLPPNSDTSKETSLT-- VNKET  107
Db  5874 VSKKIVPQKPSRTPVQEEVIEVKPVAVHTKKMVISEEKMFFASHTEEEVSVTPVEVQKEI   5933

Qy   108 TVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPT--SKVLAKPTPKAETTTK    165
Db  5934 VTEEKIHVAVSKRVE----------PPPKVPELPEKPAPEEVAPVPIPKKVPEPPAPKVPEVPK   5986

Qy   166 GPALTTPKEPTTTPKEPAS----------TTPKEPTTTIKSAPTTPKEPAP-----------   208
Db  5987 KP-VPEEKKPVPVPKKEPAAPPKVPVPEEKIPVPVAKKKEAPPAKVPEVQKGVV        6045

Qy   209 TTTKSAPTTPKEPAPTTPKEPA-PTTPKEPAPTTTKSAPTTPKEPAPTTPKK          267
Db  6046 TEEKITIVTQREESP--------PPAVPEIPKKKVPEERK-PVPRKEEEVPPPPKVPA--LPKK   6098

Qy   268 PAPTTP-KEPAPTTPKEPTPTT----------------                        288
Db  6099 PVPEEKVAVPVPVPVAKKAPPPRAEVSKKTVVEEKRFVAEEKLSFAVPQRVEVTRHEVSAEE   6158

Qy   289 ------------------------PKEPAPTTK-----------------            301
Db  6159 EWSYSEEEEGVSISVYREEEEREEEEEAEVTEYEVMEEEPEEYVVEEKLHIISKRVEAEPAE   6218

Qy   302 TT----------PKEPA--------------PTAPKK------PAPTTPKE--PAPTTPKEP-   331
Db  6219 VTERQEKKIVLKPKIPAKIEEPPPAKVPEAPKKIVPEKKVPAPVPKKEKVPPPKVPEEPK    6278

Qy   332 APTTTKEPSPTTPKE-----PAPTTTKSAPTTPKEP-- APTTTKKEAPPKARVPEEPKRAVP--TTPKEPSPTTT   382
Db  6279 KPVPEKKVPPPKVIKMEEPLPAKVTEKHMQITQEEKVLVAVTKKEAPKVPVP-IPKKLKPPPPK--VPEEPKVFEE   6338

Qy   383 KE-----------PAPTT----                                        395
Db  6339 EEKVLKLKPKREEEPPAKVTEFRKRVVKEEKVSIEAPKREPQPIKEVTIMEEKERAYTLE   6398

Qy   396 ----------------PK-EPAP----                                   402
Db  6399 EEAVSVQREEEYEEYEEYDYKEFEEYEPTEEYDQYEEYEEREYEYEEHEEYITEPEKPI    6458

Qy   403 PT--TPKEPAPTTPKEPAPTTTKPAPTAPKEPAPTAPKEPAPTTPKETAPTTPKKLTTTPEK----   456
Db  6459 PVKPVPEEPVPTKPKPAKPPPAKVLKKAVP-EEKVPVP-IPKKLKPPPPK--VPEEPKVFEE    6514

Qy   457 ---LAPTTPKERPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAP--NTPKEPAPTTPK   511
Db  6515 KIHISITKREKEQVTEP---AAKVPMKPKRVVAEEKVPVPRKEVAPPVRVPVRVPKELEPE    6571


Qy   105 KETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTK   164
Db   179 PETLVSTDSSTST-EQTSPDNTTEIASPMETNTTTEATTTSVEPSVSTLAS---EDETTV   234

Qy   165 KGPALTTPKEPTPTTPKEPASTTPKEPAPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPT   224
Db   235 TAIAEST-----TTVIAEVSTTTEEPTTT----------AESTTKKSTT---KAPA      272

Qy   225 TTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTP----KKPAPTTPKEPAPTT   280
Db   273 TTEEPTPTTTEE-VTTTEAETSTTTSSETSTEK--PTTPLIDNKIAGPATGK---PET    324

Qy   281 PKEP-TPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKE-  338
Db   325 THFPVTGTTPN--FDTATETPFVAKSEDKMTLSKTAAETTTQQTTEVT-DGPEKETTKNV   381

Qy   339 ----PSPTTPKEPAPTTPKSAP---------TTTKEPAPT---TTKSAPTTPKEPSPTTTK   383
Db   382 SIEIPITTVPLVETTSTTSTASKESDGPHTTLKLKVTTTADSDSTESATTVKPFNEETTTK   441

Qy   384 E---PAPT---------TPK-------EPAPTTPKKPAP--------------TTPK      407
Db   442 SHVPKPKTKKGTVKVTPKLELSFDEPTEIT-KAPHPGKLLEKKTYHFVLSDNFARYSEAK   500

Qy   408 E------------PAPTTPKEPAP----TTTKKPAPTA-PKEPAPTT--PKETAP        443
Db   501 ENDDYNHLDYNHYREAKEPTTTEESSTTEEVTTTEEPANTGNPPTTTENPTTTEQPTSTAE   560

Qy   444 TTPKKLTPTTPEKIA---PTTPKEPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPN   500
Db   561 STTTALPFTTEQTVTTEEPTTAEKSTATQ---KPTTTQESVST---EKTSTTKKA---S   610

Qy   501 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP   560
Db   611 TTEE--PTTTDEPTTTT--ESSTTGKATTPELSTTSEETTTTELKITTE---------GS   657

Qy   561 TTTKEPSTTSDKPAP----TTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAP    616
Db   658 TTTEEPTTTAIFAEASTGIITTDEETTSTTTSTTPEITSTKE-IVTESAITQTSVSVVES   715

Qy   617 TTP-----------KKPAPKELAPTTTKGP----------------------------TS   637
Db   716 STPRQLPERWKAIVNKFKHNLEVLKEVLKEKKRLLKEKESTSTTGSDSSETTTVVAENIDEVTT   775

Qy   638 TTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPE--TPPPTTSE-VSTPTTTKEPTTI   694
Db   776 TEKEKVVQTTPITTEKSTTQEETTTTTTTTTEKTTSKTTTEKPTTSESATTETTTSEPST-   834

Qy   695 HKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTIPE   754
Db   835 ---TEST---------------TVDTSSATTEESSTAAETTTTSAE--TSE           865

Qy   755 TTTA-------------------------------------------APKMIKETATTTEKTTE   775
Db   866 TTTSESAAFITGESPENTALQSSSQKSEENESSAEKPGARRDFVPKKHKTTVKPAETTSA   925

Qy   776 SKITATTTQ-VTST---TTQDTTPFKITTLKTTTLAPKVT---TTTKKTITTTEIMNKPEE   828
Db   926 VAASTTTEPITTTEKSTTLETTPIEATTLNEVTGPAFVTGAPVDETTINTLELLSK---   982

Qy   829 TAKPKDRATNSKATTPKPQ------------KPTKAPKKPTSTKKPKTMPRVRKP       871
Db   983 ------INNTQISQPKPTDISKTDALSLISGLIGSFTKAPMAPTI-----------    1022

Qy   872 KTTPTPRKMTSTMPELN---------PTSRIAEA                           896
Db  1023 HTTTDAAFVTATEASLNDGSDKKIIDEAQPTDEIRRA                       1059

RESULT  11
I38346
elastic titin - human (fragment)
```

7;

Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 262

```
Qy  203 PKEPAPTTKSAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKSAPTTPKEPAP 262
            :|:   ||||    |  |  |::|
Db   39 PKQTPPYYNAPPYYKSPPP----PSPSPP------PPPYVHKTPPYYKSPPP     83

Qy  263 TTPKKPAPTTPKEPTPTPKEPTTPKEPAPTTKEPAPTTP---KEPAPTAPKKPAPTT 319
            :|:    |   ||||:           ||::|
Db   84 PSPSPPPPVYKSPPPPS--PSPPPYYYKSPPPPS-PSPPPPYYYKSPPPSPSPPPPYYKSPPPPYYV 141

Qy  320 PKEPAPTTPKEPAPTTKEPSPTTPKEPAPTTPKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPSP 379
            |:|:  |  ||  |   ||                        :|
Db  142 YKSPPPPSPSPSPPPPVYKSPPPSPSPPPPYYYKSPP------------PPSPSPPPP 187

Qy  380 TTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTTTKPAPTAPKEPAPTTPK 439
            |  |   |::|   |:|:            :|:|
Db  188 YYYKSPPPSPSPSPPPPVYKSPPPSPSPSPPPPVYKSPPPSPSPPPPYYYKSPPPPSPS 247

Qy  440 ETAPTTPKKLIPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAP 499
            :|  |          :      :|||:|
Db  248 PPPPYYKSPPPPSPSPSPPPPYYYKSPPPPSPSPPPPYYYKSPPPPSPSPPPPYYYKSPPP 307

Qy  500 NTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPAPKELA 559
            |  |    ||  ||  |::|    |
Db  308 PSPSPPPPYYYKSPPPSPSPPPPYYYKSPPPSPPPPVYKSPPPSPSPPPPYYY-    366

Qy  560 PTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKKPAPTTPETPPP 619
               :|:      :|:|         |:|   |||
Db  367 ---KSPPPSPSPSPPPPYYYKSPPPPSPSPPPPYYYKSPPPPSPSPPPPYYYKSPPPPSPSPPPP 422

Qy  620 KKPAPKELAPTTKGPTSTTSDKPAPTTPKETAPTTPKKPAPTTPETPPP 677
            :|                   :|:|:  |:|
Db  423 SPP------PPYVYKSPPPPSPSPPPPYYYKSPPPSPSPPPPYYYKSPPPPSPSPPPP 475
```

RESULT 13

C84672

hypothetical protein At2g27380 [imported] - Arabidopsis thaliana

Query Match 9.6%; Score 632; DB 2; Length 761;
Best Local Similarity 31.1%; Pred. No. 2.5e-21;
Matches 217; Conservative 44; Mismatches 354; Indels 82; Gaps 33;

```
Qy  156 PTPKAETTTKGPALTTP---KEPTPT-TPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTT 211
            | |        ||:|   |||   :|
Db   69 PPPIQKPPTYSPPIYPPPIQKPPTPIYSPPIYPPPIQKPPTPTPT 128

Qy  212 KSAPTTPKEPAPTTKEPAP--TTPKEPAPTTTKSAPTTP---KEPAPT-- 263
            :  | | || :|    |            ||  |       |||
Db  129 YSPPIYP----PPIQKPPTPSYSPPVKPPPV---QMPPTPTYSPPVKPPPVHKPPTPTYS 181

Qy  264 TPKKP--APTTKEPAPTTP----KEPTP--TTPKEPAPTTKEPAPT-TPKEPAPTAPK 313
            |||    |||    |          ||:|   ||  |  |||:|
Db  182 PPIKPPVHKPPTPIYSPPIKPPVHKPPTPIYSPPIYPPPVHK 241
```

(rotated left side content)

```
Qy  512 EPA--------------------PTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 545
            |:|                          ||:|
Db 6572 EVAFEEEEVTHVEEYLVEEEEYIHEEEEFITEEEVVPIPVK-VPEVPRKVPEE-KKP 6629

Qy  546 APTTPKKPAPKELAPTTTKEPTSTT------SDKPAPTTPKGTAPTTPKEPAPTTPKEPA 599
            ||  |   | :   |:|          ||
Db 6630 VPVPKKKEAPPAKVPEVPKKPEEKVPVLIPKKEKP----PPAKVPEVPKKKPVP-EEKVPV 6684

Qy  600 PTTPKGTAPTTLKEPA--PTTPKKPAPKEL----APTTKGPTSTTSDKPAPTTPKETAP 653
            |||           :|: :                ||:|
Db 6685 PVPKKVEAP-----PAKVPEVPKKPVPEVPAPKKVEAPPAKVPEVPKKLIPEEKKP 6739

Qy  654 T-TPKE--PAPTTPKKPAPTTPETPPPTSEVSTPTTTKEPTTIHKSPDESTPELSAEP 709
            |  |||    |||  |                :|    :|
Db 6740 TPVPKKVEAPPPKVPKKREPVPVPVALEQEEEVLFEEEIVPEEEEVLPE-EEEV 6798

Qy  710 TPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDL--RTTPETTTAAPKMTKETA 767
            |           |:|::          |:                 :|
Db 6799 LPEEEEVLPEEEEIPEEEEVPEEEEYVPEEEEVLPEVKPKVPVPAP------VP 6853

Qy  768 TTTEKTTESKITATTTQVISTTTQDTTPFKITTLKTTLAPKVTTTKKTITTTEIMNKPE 827
            |||:           |::           |  ||           :|
Db 6854 EIKKKVTEKKVVIPKKEEAPPAKVPEVPKKVE--EKRIILPK----EEEVLPEVTEEPE 6907

Qy  828 ETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKK-----PKTMPRVRK---PKT-TPTPR 878
            :|                 :|::| |:         :|   ::|
Db 6908 E-----------EPISEEEIPEEPSIEEVEEVAPPRVPEVIKKAVPEAPTPVPK 6951

Qy  879 KMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHV 938
            |                         :|                        |||:
Db 6952 KVEA--PPAKVSKKIPEEKVPVPVQKKEAPPAKVPEV--------------------- 6986

Qy  939 FMPEVIPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRNGTLVAFR 991
            |                                            ||||
Db 6987 P--PKKVPEKKVL--VPKKEAV-----------------------PPAKGRTVLEEKVSVAFR 7021
```

RESULT 12

T11622

extensin class 1 precursor - cowpea

Query Match 9.68%; Score 633; DB 2; Length 489;
Best Local Similarity 31.2%; Pred. No. 1.5e-21;

```
Qy   314 KPAPT-TPKEPAPTTPKEPAP---TTPKEPSPTTPKEPAPTTKSAPTTTKEPAPTTKS 369
         |::||  |  |:| |  ::    || |  :|| | : |       ||  :| ||
Db   242 PPTPIYSPPIKPPPVHKPPTPIYSPPVKPPPVQTPPTPIYSPPVKPPPVHKPPTPYSPP 301

Qy   370 APTTPKEPSPTTTKEP---APTTPKEPAPT--TPKKPAPTTPKEPAPTTPKEPAPTTKK 424
         | |  |  :| |      | | |  | |     |  || |  | |  |  | |
Db   302 VKSPPVQKPPTPTYSPPIKPPPVQKPPTPIYSPPVKP-PPVHKP-PPVHKP 360

Qy   425 PAP--TAPKEPAPT---TPKETAPTTP---KKLLPTTPEKLAPTTPEKPAPTT---PE 471
         | |   |  | |      |  | |      ::| |   | |  | | | | |
Db   361 PTPIYSPPVKPPPVHKPPTPIYSPPVKPPPIQKPPTPIYSPPIKPPPLQKPPTPYSPPI 420

Qy   472 ELAPTTPEEPTPTTPEEPAPT---TPKAAAP-NTPKEPAPTTPKEPAPTTPKEPAPTTP 526
          | |    |  |      |      |      || | |  | |  | |  |  | |
Db   421 KLPPVKPPTPIYSPPVKPPPVHKPPTPIYSPPVKPPPVHKPPTPIYSPPFIKPPPVKPPTP 480

Qy   527 KETAPTTPKGTAPTTLKEPAPT--TPKKPAPKELAPTTKEPTSTTSDKPAPTTPKGTAP 584
          :| |  | | |    | |      |  ||   | |     |:||:| | |  | ||
Db   481 TYSPPVQP---PPVQKPPTPTYSPPIKPPPIQKPPT----PTYSPPIKPPPVKPPTPTY 532

Qy   585 TTPKEPAPTTPKEPAPT-TPKGTAPTTLKEPAPT--TPKKPAPKELAPTTKGPTSTTSD 641
         : | |  | |  | |    | | |    | |      |  ||   | |     |:|  |
Db   533 SPPIKP-PPVHKPPTPTYSPPIKPPPIHKPPTPTYSPPVKPPPVHKPPTPTYSPPIKPPP 591

Qy   642 KPAPTTPKETAPTTPKEPAPTTPKKPAPT-TPETPPTTSEVSTPTTTK--EPTTIHKSP 698
          |  |  |:| |  |  | |  |  ||    | | |:|  |  | :|    |||:| |
Db   592 VHKPPTPTYSPPIKP----PPVHKPPTPTYSPPIKPPPVHKPPTPTYSPPIKPPPVHKPP 647

Qy   699 DESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTA 758
           ||    | |   |   | |  |  | | |   |         |       |  ||:|
Db   648 ---TPTYSPPIKPPPVQKPPTPTYSPPVKPPPVQLPP-TPTYSPPVKPPPVQVPPTPTYS 703

Qy   759 APKMTKETATTTEKTESKITATTTQVTSTTTQDTTP 795
          |   |  |    |                | |||
Db   704 PPVKPPPVQVPPTPTYSPPIKPPPVQVPPTPTPSPP 740


RESULT  14
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34513
R;Favello, A.; Vaudin, M.
A;Description: The sequence of C. elegans cosmid ZK783.
submitted to the EMBL Data Library, August 1994
A;Reference number: 221536
A;Accession: T34513
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3507 <FAV>
A;Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A;Experimental source: strain Bristol N2; clone ZK783
C;Genetics:
A;Gene: CESP: ZK783.1
A;Map position: 3
A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3304/1

Query Match    9.6%;   Score 629;  DB 2;  Length 3507;
Best Local Similarity  24.1%;  Pred. No. 1.6e-20;
Matches 252;  Conservative 158;  Mismatches 416;  Indels 218;  Gaps 45;

Qy    34 KKKPTPKPPVVDEAGSGLDN-GDFKVTT----PDTSTTQHNKVSTSPKITTA------K 81
         ||  |   |               |  |      |::|| |  |:|| |:|:      |
Db  1942 KKQPNRKIEIDEENSSSNSGQEKPTTKGIVSSVTSATSSESTTAEPHVTTISSSTTSTK 2001

Qy    82 PINPRPSLPPN---SDTSKETSLITVNKETTV-ETKETTTNKQTSTDGKEKTTSAKETQS 137
          :|  | |     :||| |:| | | | |||  ::|   ||| || :   |||:|||:
Db  2002 DMTSSKS-PENVTMSSESPEVSTSSSKSTTASETTVSSTPSESSSSEAPLTSSPATTTEV 2060

Qy   138 IEKTSAKDLAP----TSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTT 193
          ||:| |          |||  ||  |||||| || ||||||||||||||||:||||| |
Db  2061 ITESSVKSTPKEESSSEITVKLSSKSPEVTESSVKSSPSTPS-TTSQSVTSTVPETSKS 2119

Qy   194 TTIKS-APTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTKEPAPTTTKS 252
         || | ||| |  |:| |:||  || | |    |  | | |  ||||:|
Db  2120 TVLSSEAPVTSTSPTEVHT-SSETKPSLSASSTGDTNSTTPSTSLASVKSTSAPEGTS 2178

Qy   253 APTTPKEPAPTTPKPAPTTPK-EPAPTTPKEPPTTPKEPAPTTKEP-APTTPKEPAPT 310
         |   |::| | | :  ||    :  | |     | |        :|  | |  | |
Db  2179 ASVAPVKLSSLSPDVSQPSTKTFDATESSTVQASETSSGTSVKSTSEPESHVTKLSITSS 2238

Qy   311 APKKPAPTTPKEPAPTTPKEPAPTTKEPSPTTP--KEPAPTTTKSAPTTTKEPAPTTTK 368
          |  | |  |  :  | |    | |   |: |    :  |  :|    |  || | |
Db  2239 NPSSSVPVTSPKSTPTVPE----STEQPTSTTPSGQSLTPMNSNSEVLTTSEPHVLSSS 2293

Qy   369 SAP------TTPKEPSPTTKEPAPTTPKKPAPTTPKEPAPTT-PKEPAPTT- 421
         |       || | |:| ||  | | |  | |  | |  | |  | | | |
Db  2294 LSPDVSQSSTTPNNLSESSTVE---TPKTSSEVSLNSEEPSTTEAPTTLSPDILSTTTN 2349

Qy   422 ------------------TKKPAPTAPK-------EPAPTTPKETAPTTPKKLTP 451
                           ::| | | |        | |  | |:| |  |:|
Db  2350 NLSQSSTVSTEDRSEISSENSEKPT-SAPELVTSSVTHVASSSPDVPTES--SEPDDLTG 2406

Qy   452 TTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEP----------APTTPKAAAPN 500
         :|  |:|  | |  | |  | |  |  |  | |  ::           | |  |
Db  2407 SSTENIPEASSKQTISSTTPDTTASEEPTKSTSMSPDLSTTSNVLSESSTTPESSS- 2464

Qy   501 TPKEPAPTT------------PKEPAPT----TPKEPAP------TTPKETAPTTP 534
          |:  |               | | |       | |           |:| |
Db  2465 --KSPVSSSTEGISVVTSTEFSKVPESTISSVLEEDLTKTTPSPILEETTTASETSEPLT 2522

Qy   535 KGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPA---------PT-------- 577
          ||  | |  | |:| :  |            | ||:|  |          |
Db  2523 EDSLTVSVRIHELTTSSENVPKESESTTTSSESSKPSQEPAGILTSTVVPTSSVSLITA 2582

Qy   578 --------TP--KGTAP--TTPK---------EPAPTTPKEPAPTTPKGTAPTTLKEPAPT 617
                 |    | | |   |:|          | | | |  | |  | | |    | |
Db  2583 SEIEAITSNTPFKQGRTPITTSPKSLVKSTTSPSTVVTSSESTKRTTVSTTVSTTTPT 2642

Qy   618 TPKKPAPKELAPTTKGPT-STTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTPETPP 676
          |  ||   | |     | :|:| | | |  |:| |  |  | |  |  || | | |
Db  2643 EETTTSESLILTAAPSKPTESTTESSEAPTTPAKTSETKPSNVSSTSRKSTENVETSTSQ 2702

Qy   677 PTTSEVSTPTTTKEPTTIHKSPDESTP-ELSAEPTTPKALE-NSPKEPGVPTTKTPAAT 732
          || |     |  |       | ||    |  | | |     :||  |  | | | |
Db  2703 SGSLESSTMSSTS------SEPETNAPATVSSEASSTTLEENSSTSSP---TSSEASV 2752

Qy   733 K----PEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTESKITATTT---- 783
              | | | |  |:|    | |      | |||  |  ||  | |    |:|
Db  2753 KLSSLEPESITSEAVTVSSR---APAEITMSSESHREISTVSSEPSEPEIPLSTTVSPN 2808

Qy   784 ----------QVTSTTQDTTP-FKITT------LKTTTLAPKVTTTKKTITTTEIMNKP 826
                    |:|| | | |  |:|       | ||:| | |||:| :|||:|:   |
Db  2809 VVTASSIPSEEPILSSVTSSSTPRVRLITGTPDDLLVSVTVPSHGNRRQNITASSV---P 2865

Qy   827 EETAKPKDRATNSKATTPKEQKPTKAPKKPTSTKPKTMPRVRKPK--TTPTPRKMTST 883
         | | |   |    |  || | |  | |||  |||| :|  |||    |:|:|||||:|
Db  2866 SNSTSPIILPSES-LTTPQPPPTTTTAKPATTSGKRGPPSIQPPAEMFTTPAP------ 2918

Qy   884 MPELN------PTSRIAEAMLQTTT 902
          |          || |||
Db  2919 PPPSNGGYGEETNQEEQVTSTTT 2942


RESULT  15
I51618
nucleolar phosphoprotein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C;Accession: I51618; S57757
R;Cairns, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A;Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp
```

Mon Apr 29 08:35:14 2002

aal.rpr

```
A;Reference number: I51618; MUID:96019267
A;Accession: I51618
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-990 <CAI>
A;Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921
C;Genetics:
A;Gene: xNopp180
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

Query Match         9.5%;  Score 627; DB 2;  Length 990;
Best Local Similarity   26.2%;  Pred. No. 5.5e-21;
Matches 258;  Conservative 134;  Mismatches 401;  Indels 192;  Gaps  46;

Qy       7 PIYLLLLSVFVIQQVSSVKDNKKNRTKKKPTTPKPPVVDEAGSGLDNGDFKVTTPDTSTT  66
           | ||:   |  ||     :  ::   | |  | ::  ||   : ::  | :: |
Db      44 PTSLLDIFSDWV----KSPDAKKRKPPANGLPKKKSAKESSSSEDSSSEEDEPPAKKRA  98

Qy      67 QHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVE--TKETTTTNK----Q 120
           |  :   |:   | | :  ::    : :  || ||  |: |: |  : :
Db      99 Q----PAGGKKPVVKAVQPKKAKSSSEDSSDESD--SEEETKKPPAKRPAQTPKVAAVK 151

Qy     121 TSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTP 180
           |  | | || : :| | ::   |  | : |      ||     | |
Db     152 TPTQKKAKSSSS-ESSSSEDEASKKKQPVIKV----PKQAVVKAGLASNNGKTADSSSS 206

Qy     181 KE----PASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTPKE 236
           ||      |   |  |    :   |: :   | |  :  ||   | |   :  ||||
Db     207 EDSDSPPAKKTAAIKTPPT---KPATAAKPQAKKTAGKKSSSREDSSDSSDEEQKTAKSK 263

Qy     237 PAPTT-TKEPAPTT-----TKSAPTTPKEPAPT---------TPKKPAPTTPKEPAPTT 280
           ||||   |||| |      |  | || ||||            | ::|  || ||||||
Db     264 PKPDVSAVPPPTSVSKKKTLSQPGTKAKPESSDSDSDSDEEEQPAKKAKIVPAKAAASA 323

Qy     281 P----KEPTPTTPKEPAPTTKEPAPTTPKPAPTTPKEPAPTTP----KEPA 332
           |       |  |   :  |:       |  |  |  ||     :      | |
Db     324 PKPLAKKAETSTDSEDSDSSSEDEKKSSVKLGVKAAPKK-APAAP--DAKSTPVAAAKKSA 380

Qy     333 P---------------TTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPK 375
           |               |||    :  || |     :     || | ::|||||
Db     381 PAKKASSSSDSDSSSNEETTTKPAAKTTPAKSAATTPSKTPTNGK--ATPTSKTPAKPG 437

Qy     376 EPSPTTTKEPAPTTPKEPA-----PTTPKKPAPTTPKEPAPT-TPKEPAPTTTKKPAPTAP 430
           |  | ||  ||  ||        | ||  | |     ||     |  | ||  ||| |
Db     438 TPKTSTAKKDSSSSDSSDSSEDEETTTKPAAKTTPAKSAATTPSKTP---TNSKATPTSK 494

Qy     431 KEPA-PTTPKETA---------PTTPKKLIPTTPKEPAPTTTPEKPAPTTPEELAPTT 477
           | |: || ||          ||   :  || | |   | ||| ||     | |||
Db     495 KTPAKPGTPKITSAAKKDSSSSDSSSSSDSSSDEKKTPA--KRAAKTTPAKPA------AKTT 545

Qy     478 PEEPTPTTPEEPAPTTP-KAAAPNTPKEPAPT----TPKEPAPTTPKEPAPTTPKETAPT 532
           |  |  ||  | |    |:  :   | ||:       |  ||| ||  | ||  | |  |
Db     546 PAKPA------AKTTPAKPAAKSTPGKQVPTKKESSSSDSDSSDSSEDEKKSSAKPAVKT 598

Qy     533 TP-KGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPA 591
           || |  ||    | | | ||  | |   ||       |  | ||     |||||  ||| :
Db     599 TPGKATS------KPVVASKPVPAK------KASSSSDSSSEEETTKITKPLIKLSPA 645

Qy     592 PTT--PKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGP------TSTTSD 641
           |||  | || ||   |   | |    |   ||                  :  |||||
Db     646 VKTLPPKKAESSSDSSSDSDSEKK---TKPAKPPAKSATPVNTKAPAQNKASKASCSDSD 702

Qy     642 KPAPTTPKETAPT--TPKEPAPTTPKK-PAPTTPETPPPTTSEVSTPTTTKE-PTTIHKS 697
           ||      | |    |||| |  | ::  |   |:       ::  || :|  |||
Db     703 SSEEEGKSKQPTGKSPAAKATAPPKKNPVAYNKDKPSSSSSDSSGCDDEKQPKQAAAA 762

Qy     698 PDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPETTT 757
           ||     |          : :::::||     |       ||||  :   | |||||||
Db     763 KDVKQGAKAAKPTPKKAASSSE---DSSSDEDVSKAKKTNTAVSKSPV----TTPKAVP 815

Qy     758 AAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKIITTLKTTTLAPKVTTTK--K 815
```

```
Db     816 AAKKESSESSSDSEDEKQGGKNTSTTKIANST----------------PKAAAAECSE 857
           ||| :      :: :|                                      ||
Qy     816 TITTTEIMNKPEETA-KPKDRATNS---KATTPKPQKPTKAPKKPTSTKKPKTMPRVRKP 871
           :::|   | :   | :|   |   ::  | :| : | ::
Db     858 ESSSSEDEGKANGTSGKRKRESTGNAECEAVTPE-------NKKLKAKSPNTFPKVNKK 909

Qy     872 KTTPTP-RKMTSTMPELNPTSRIAE 895
           :  ||  |  :::  ::|
Db     910 ELKNTPFRRVVEDIEINP--RMAD 932
```

Search completed: April 26, 2002, 16:30:58
Job time: 653 sec
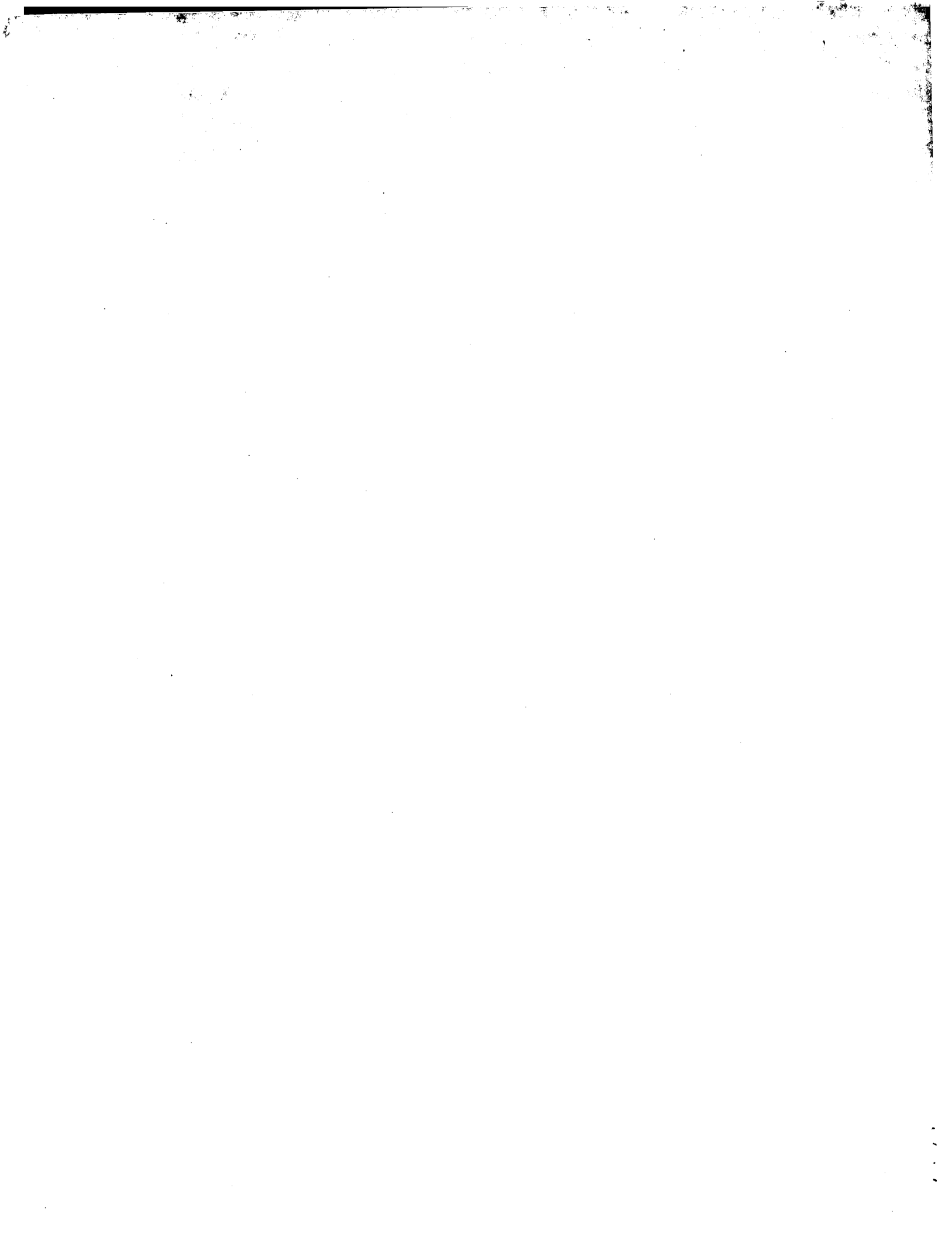
aa1.rpr

```
           GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     April 26, 2002, 16:22:50 ; Search time 62.75 Seconds
                                         (without alignments)
                                         718.105 Million cell updates/sec

Title:           AA1
Perfect score:   6568
Sequence:        1 MAWKTLPIYLLLLSVFVIQ.........ARAITTRSGQTLSKVVWNCP 1229

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters:          100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_39:*

   Pred. No. is the number of results predicted by chance to have a
   score greater than or equal to the score of the result being printed,
   and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	Query Match %	Length	DB	ID		Description
1	1168	17.8	5179	1	MUC2_HUMAN	Q02817	homo sapien
2	950	14.5	1664	1	SLP1_CLOTM	Q06852	clostridium
3	789.5	12.0	1367	1	AMYH_YEAST	P08640	saccharomyc
4	651	9.9	1255	1	MUC1_HUMAN	P15941	h mucin 1 p
5	556.5	8.5	875	1	FP1_MYTED	Q25460	mytilus edu
6	555.5	8.5	2700	1	ZAN_HUMAN	Q9y493	homo sapien
7	551	8.4	1087	1	EXTN_TOBAC	P13983	nicotiana t
8	533	8.1	620	1	NFH_MOUSE	P19246	mus musculu
9	530.5	8.1	1162	1	TCNA_TRYCR	P23253	trypanosoma
10	530	8.1	865	1	CPN_DROME	Q02910	drosophila
11	519	7.9	872	1	FP1_MYTCO	Q25434	mytilus cor
12	509.5	7.8	662	1	MUC1_XENLA	Q05049	xenopus lae
13	503.5	7.7	1970	1	RPB1_HUMAN	P24928	homo sapien
14	498.5	7.6	1970	1	RPB1_MOUSE	P08775	mus musculu
15	497	7.6	831	1	NFH_RAT	P16884	rattus norv
16	493.5	7.5	467	1	RPB1_CRIGR	P11414	cricetulus
17	488.5	7.4	826	1	SSP2_PLAYO	Q01443	plasmodium
18	488	7.4	1020	1	NFH_HUMAN	P12036	homo sapien
19	475.5	7.2	267	1	EXTN_MAIZE	P14918	zea mays (m
20	471.5	7.2	5376	1	ZAN_MOUSE	O88799	mus musculu
21	468.5	7.1	634	1	HWP1_CANAL	P46593	candida alb
22	467	7.1	817	1	VRP1_YEAST	P37370	saccharomyc
23	463	7.0	2142	1	BAT2_HUMAN	P48634	homo sapien
24	459	7.0	1161	1	YJ9P_YEAST	P47179	saccharomyc
25	454	6.9	797	1	VGLX_HSVEB	P28968	equine herp
26	448.5	6.8	670	1	VG50_HSVI1	Q00130	ictalurid h
27	442.5	6.7	751	1	FP1_MYTGA	Q27409	mytilus gal
28	439.5	6.7	1083	1	T2D3_HUMAN	O00268	homo sapien
29	432.5	6.6	439	1	XP2_XENLA	P17437	xenopus lae
30	432	6.6	3164	1	TEGU_HSV11	P10220	herpes simp
31	426.5	6.5	1794	1	YAV1_SCHPO	Q10172	schizosacch
32	426	6.5	2715	1	TRX2_HUMAN	Q9um6	homo sapien
33	424.5	6.5	1125	1	MAP4_MOUSE	P27546	mus musculu
34	424	6.5	307	1	SGS3_DROME	P02840	drosophila
35	424	6.5	1251	1	YQU3_CAEEL	Q09550	caenorhabdi
36	422.5	6.4	3421	1	TEGU_HSVEB	P28955	equine herp
37	421	6.4	1229	1	N121_HUMAN	Q9y2n3	homo sapien
38	419.5	6.4	2476	1	ZAN_PIG	Q28983	sus scrofa
39	419.5	6.4	2774	1	MAPA_RAT	P34926	rattus norv
40	419	6.4	1185	1	DRPL_HUMAN	P54259	homo sapien
41	411.5	6.3	907	1	VGP3_EBV	P03200	epstein-bar
42	410	6.2	1183	1	DRPL_RAT	P54258	rattus norv
43	410	6.2	1411	1	TCOF_HUMAN	Q13428	homo sapien
44	407.5	6.2	3256	1	KI67_HUMAN	P46013	homo sapien
45	405.5	6.2	2517	1	NCR2_HUMAN	Q9y618	h nuclear r

```
                        ALIGNMENTS

RESULT   1
MUC2_HUMAN
ID   MUC2_HUMAN     STANDARD;      PRT;   5179 AA.
AC   Q02817; Q14878;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
GN   MUC2 OR SMUC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=94132002; PubMed=8300571;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT   "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT   Identification of the amino terminus and overall sequence similarity
RT   to prepro-von Willebrand factor.";
RL   J. Biol. Chem. 269:2440-2446(1994).
RN   [2]
RP   SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=93016075; PubMed=1400449;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA   Kim Y.S.;
RT   "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT   both upstream and downstream of its central repetitive region.";
RL   J. Biol. Chem. 267:21375-21383(1992).
RN   [3]
RP   SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX   MEDLINE=91358717; PubMed=1885763;
RA   Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA   Petersen G.M., Kim Y.S.;
RT   "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT   and polymorphism.";
RL   J. Clin. Invest. 88:1005-1013(1991).
CC   -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC       OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC       PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC       AGENTS AT MUCOSAL SURFACES.
CC   -!- SUBUNIT: MULTIMERIC.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC       BRONCHUS, CERVIX AND GALL BLADDER.
CC   -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC       INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC       VARIES AMONG DIFFERENT ALLELES.
CC   -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC       OF SILKWORM HEMOCYTIN.
CC   -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
```

```
DR   EMBL; L21998; AAB95295.1; -.
DR   EMBL; M74027; AAA59875.1; -.
DR   EMBL; M94131; AAA59163.1; -.
DR   EMBL; M94132; AAA59164.1; -.
DR   MIM; 158370; -.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002400; GF_cysknot.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF00094; vwd; 4.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   SMART; SM00214; VWC; 2.
DR   SMART; SM00011; VWC_def; 2.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 2.
KW   Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   5179       MUCIN 2.
FT   DOMAIN     1747   1747       APPROXIMATE REPEATS.
FT   REPEAT     1401   1416       1.
FT   REPEAT     1417   1432       2.
FT   REPEAT     1433   1448       3.
FT   REPEAT     1449   1464       4.
FT   REPEAT     1465   1471       5.
FT   REPEAT     1472   1478       6.
FT   REPEAT     1479   1494       7A.
FT   REPEAT     1495   1517       7B.
FT   REPEAT     1518   1533       8A.
FT   REPEAT     1534   1556       8B.
FT   REPEAT     1557   1572       9A.
FT   REPEAT     1573   1596       9B.
FT   REPEAT     1597   1612       10A.
FT   REPEAT     1613   1635       10B.
FT   REPEAT     1636   1651       11A.
FT   REPEAT     1652   1675       11B.
FT   REPEAT     1676   1683       12.
FT   REPEAT     1684   1699       13.
FT   REPEAT     1700   1715       14.
FT   REPEAT     1716   1731       15.
FT   REPEAT     1732   1747       16.
FT   DOMAIN     4815   4886       VWFC 1.
FT   DOMAIN     4924   4991       VWFC 2.
FT   DISULFID   5075   5160       CTCK.
FT   DISULFID   5089   5122       BY SIMILARITY.
FT   DISULFID   5098   5136       BY SIMILARITY.
FT   DISULFID   5102   5154       BY SIMILARITY.
FT   DISULFID      ?   5159       BY SIMILARITY.
FT   CARBOHYD    163    163       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    670    670       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    770    770       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    894    894       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1139   1139       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1154   1154       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1215   1215       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1230   1230       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1246   1246       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1787   1787       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1820   1820       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4339   4339       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4351   4351       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4362   4362       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4373   4373       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4422   4422       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4438   4438       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4502   4502       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4616   4616       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4627   4627       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4752   4752       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4787   4787       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4881   4881       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4888   4888       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4955   4955       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   4970   4970       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   5019   5019       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   5038   5038       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   5069   5069       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT   1351   1351       H -> L (IN REF. 3).
FT   CONFLICT   1412   1412       T -> S (IN REF. 3).
FT   CONFLICT   1449   1449       L -> P (IN REF. 3).
FT   CONFLICT   1504   1504       M -> T (IN REF. 3).
FT   CONFLICT   4192   4192       G -> S (IN REF. 2).
SQ   SEQUENCE   5179 AA;  540295 MW;  85CD7571FB9A5663 CRC64;
```

```
Query Match              17.8%;  Score 1168;  DB 1;  Length 5179;
Best Local Similarity    32.5%;  Pred. No. 9.1e-41;
Matches  363;  Conservative  53;  Mismatches 448;  Indels 254;  Gaps 40;

Qy     45 DEAGSGLDNGDFK-----------------------VTTPDTSTTQH-NKVSTSPK------- 76
          |  ||::||| :|                       | || | ||  || | |:
Db   1312 DHPSSGSDDGDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKYQCDVSVGFICKNEDQF 1371

Qy     77 ------------------ITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETK 112
                            |||| | | || ::|:    || |  | :
Db   1372 GNGPFGLCYDYKIRVNCCWPMDKCITTPSPPTTTTTLPPTTTPSPPTT---- 1427

Qy    113 ETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTP 172
          | | ::|                          |::  |||   ||| | ||
Db   1428 -TTTTPPTTTPSPPITTTTTP----------LPTT----TPSPPISTTTPPPTTTP 1470

Qy    173 KEPT-------PTTPKEPASTTPKEPTPTTIKSAP-TTPKEP-------APTTTKSAP-TTP 218
          ||||       ||:   | | | ||:|     |||:|       |||  ||  |||
Db   1471 SPPTTTPSPPTTTTTPSPPTTTTTPPTTTPTPPASTTTLPPTTTPSPPTTTT 1530

Qy    219 KEPAPTTTKSAP-----TTPKEPAPTTTKSAP----TTPKEPAPTTPKKPA 269
          |||||  ||        |||  |:        |||  ||||
Db   1531 TTPPTTTPSPPTTTPITPPTSTTLPPTTTPSPPTTTTPPTTTPSPPT 1590

Qy    270 PTTPKEPAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEP------APTAPKKPAPTTPKE 322
          ||||  |:    | | | ||:|          ||||    |||  |||
Db   1591 ITTTTPPTTTPSPPTTTPPTTTPSPPTTTPITPPTSTTLPPTTTPSPPTTTTT 1650

Qy    323 PAPTTKEPAPTTTKKEPSPTTPKEPAPTTTKSAPTTTKSAP---TTPKEPSP- 379
          |  ||| | ||:  |||   ||||  |:      |||   |||:
Db   1651 PPPTTTPSPPTTTTTPSPPITTTTTPPTTTTTPSPPTTTTPSSPI- 1709

Qy    380 TTTKEPAPTTPKKPAPTTPKKKAPTTPKEPAPTAPKEPAPTTP- 438
          |||  |  |||  || |||  |:  ||||  ||:
Db   1710 TTTTTPSSTTTPSPPTTMTTTPSPPTTMTTLPPTTTSSPLTTTLPPSITPPT 1769

Qy    439 ---KETAPTTP------------------------KKLTP 451
             ||:||||                        ;
Db   1770 FSPFSTTTPTTTPCVPLCNWTGWLDSGKPNFHKPGGDTELIGDVCGPGWAANISCRATMYP 1829

Qy    452 ---------------------------TTPEKLAPTT 461
                                     |
Db   1830 DVPIGLGQTVVCDSVGLICKNEDQKPGGVIPMAFCLNYEINVQCCECVTQPTTMTTTT 1889

Qy    462 PEKPAPTTPEELAPTTPEEPTPT-TPEEPAPTTPKAAAPNTPKEPAPTTPKEP------AP 515
          ||||||||||  |||  |  |||:             |||   |||
Db   1890 TENPTPPTTPITTTTTVIPTTPTPGTQTPTTTTTTTVTPTTPTPGTQPTTTTPITT 1949
```

```
KW   Cell wall; s-layer; Signal; Repeat.
FT   SIGNAL        29     28                    POTENTIAL.
FT   CHAIN         29   1664                    CELL SURFACE GLYCOPROTEIN 1.
FT   DOMAIN        36    763                    4 X 156 AA APPROXIMATE REPEATS.
FT   REPEAT        36    191                    1.
FT   REPEAT       207    363                    2.
FT   REPEAT       409    565                    3.
FT   REPEAT       607    763                    4.
FT   DOMAIN       771   1377                    APPROXIMATE TANDEM REPEATS OF
FT                                              T-P-S-D-E-P.
FT   DOMAIN      1378   1449                    GLY/PRO/SER/THR-RICH.
FT   DOMAIN      1453   1494                    SLH 1 (INCOMPLETE).
FT   DOMAIN      1495   1565                    SLH 2.
FT   DOMAIN      1566   1625                    SLH 3.
FT   DOMAIN      1626   1646                    SLH 4 (INCOMPLETE).
SQ   SEQUENCE   1664 AA;  178194 MW;  5F39669SBA9FE74B CRC64;
```

```
Query Match                14.5%;  Score 950;  DB 1;  Length 1664;
Best Local Similarity      31.0%;  Pred. No. 2.7e-32;
Matches 312;  Conservative 104;  Mismatches 361;  Indels 230;  Gaps 55;
```

```
Qy     152  VLAKPTP-KAETTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTT  210
              :  : ||  :  |:   ||  ||:: |||:: |   ||  ||||            ||
Db     758  VVIQAPIKAASDEPIPTDTPSDEPTPS---------DEPTPS--DEPTPSDEPTPSD  804

Qy     211  TKSAPTTPKEPAPTTKEPAPTTKEPA----PTTTKEPAPTTTKSAPTTPKEPAPTTPK  266
              ||   |  |  ||  |   ||        ||   ||   ||    ||:|: |||
Db     805  EPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTP  864

Qy     267  KPAPTTPKEPAPTTPKEPITPTTPKEPAPAPTTKEPAPT-TPKEPAPTTAPKKPAPTTPKEPAP  325
              ||||       |  ||    ||   ||:|::: |||     || |||| :|   |
Db     865  SDEPTPSDEPTPS--DEPTPS--DEPTP-SDEPTPSETPEEPIPTDTPSDEPTPSDEPTP  919

Qy     326  TTPKEPA--PTTTKEPSPT-TPKEPAPTTTKS-APTTKEPAPT---TTKSAPTTPKEPS  378
              || ||:    | |  :|  ||||||    ||  |||  ||||   |:  |||| |
Db     920  SDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPT  979

Qy     379  ----PTTTKEPAPT-TPKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTTPKEPAPTT  433
                  || |||    ||  ||||   || ||   ||  ||      |||   ||
Db     980  PSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEP  1039

Qy     434  APT---TPKETAPTTPKKLTPTTPKELAPTTPKKPAPT---TPEELAPTTPEEPTPT-TP  486
              |: |     ||:|   :||  |  |  |||||   ||:|        || |::  ::  |:
Db     1040 TPSDEPTPSE----TPEEPIPTDTPSDEPTPSDEPTPSD-EPTPSDEPTPSETP  1094

Qy     487  EEPAPTTPKAAAPNTPKEPAPTTPKEPA--PTTPKEPAPT-TPKETAPTTPKGTAPTTLK  543
              ||  ||||   |  ||  ||     |     ||||||    |||  ||||    ||
Db     1095 EEPIPTDTPSDEPTPSDEPTPS---DEPTPS--DEPTPSETPEEPIPTDTPSDEPTPSD  1154

Qy     544  EAPATTPKKPAPKELAPTTTKEPT-STTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTT  602
              |  ||||     ||:|   |:| |    ::||   |    | ||| ||  || ||
Db     1155 EPTPS--DEPTPSD-EPTPSDEPTPSETPEEPIPTDTPSDEPTPS--DEPTPS-  1208

Qy     603  PKGTAPTTLKEPAPT-TPKKPAPKELAPTTKGPTSTTSDKPAPT---TPKETAPTTPKE  658
              ||   ||  ||||     ||| |:      |  |: ||     ||    |||  |||
Db     1209 ---DEPTPSDEPTPSETPEEPI----------PTDTPSDEPTPSD-EPTPSDE  1253

Qy     659  PAPTTPKKPAPT-TPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENS  717
              || ||||||     |:|||| :  |||  ||||   |    | || || ||| |
Db     1254 PTPS-DEPTPSETPEEPIPTDTPSDEPTPSDEPT----PSDEPTP--SDEPTPSDEPTP  1305

Qy     718  PKEPGVPTTKTPAAT--KPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTE  775
              ||||  ||  |      |||| |:|| |: || || |   |:||:|:|   ||||||
Db     1306 SDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSD----EPTPSDEPTPSDEPTPS  1361

Qy     776  SKITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKTTTTTEIMNKPEETAKPKDR  835
              ||                 |:       |:  ||||| ||  |    |    |
Db     1362 DEPTPSDEPTPSETPEEPTP---------TTTTPTTPSTT------PTSG  1396

Qy     836  ATNSKAT------TPKPQKPTKAP-KKPTSTKKPKTMPRVRKPKTTPTPRKM------  880
              |             |  | | ||   |:   |     ::  | |
Db     1397 SGGSGGGGGGGGGTVPTSPTPTPTSKPTSTPAP---TEIEEPTPSDVPGAIGGEHRA  1453
```

```
RESULT   2
SLP1_CLOTM
ID   SLP1_CLOTM     STANDARD;      PRT;   1664 AA.
AC   Q06852;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
DE   PROTEIN 1).
GN   OLPB.
OS   Clostridium thermocellum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1515;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIB 10682;
RX   MEDLINE=93209931; PubMed=8458832;
RA   Fujino T., Beguin P., Aubert J.-P.;
RT   "Organization of a Clostridium thermocellum gene cluster encoding the
RT   cellulosomal scaffolding protein CipA and a protein possibly involved
RT   in attachment of the cellulosome to the cell surface.";
RL   J. Bacteriol. 175:1891-1899(1993).
CC   -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC   -!- SUBCELLULAR LOCATION: CELL WALL.
CC   -!- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X67506; CAA47841.1; -.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 3.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
```

```
Qy     516  TTPKEPAPT-TPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKEPTSTTSDKP  574
              || |||     |  |  ||| ||||   |||      ||||||| |||::    |
Db     1950 TTTVTPTPTGTQTPPTPITTTTVPTPTPTGTQTPPTPTGTQ  2009

Qy     575  AP-TTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTTK  633
              |    |||  ||       ||  |  ||| || |      ||| ||||||| |:|| :|
Db     2010 TPTTTPITTTTVTPTPTPTGTQTPT--TPITTTTVTPTPTPTGTQTPPTPITTTTTV  2068

Qy     634  GPTSTTSDKPAP-TTPKETAPTTPKEPAPTTPETPPPTSEVSTPTTTKEPT  692
              ||:: ::||      ||  |  |||| |  |   |    |: ||:  |:
Db     2069 TPTTTPTGTQTPTTPITTTTVTPTPTPTGTQTPT--TPITTTTTVTPTPTPTGTQTPT  2127

Qy     693  TIHKSPDESTPELSAEPTPKALENSPKEP----------GVPT-TKTPAATKPEMTTTAK  741
              |    |||| ||  |::  ||            ||:  |||| |     |
Db     2128 T----TPITTTTVTPTPTPTGTQTPTTTPITTTTVTPTPTPTGTQTPPTPITTTTTVT  2184

Qy     742  DKTTERDLR---TTP--ETTTAAPKMT---KETATTTEKTTESKITATTTQVTSTTTQDT  793
              ||:   ||    |||    | |             :|  :||   |  || |   |
Db     2185 PTPTPTGTQTPPTPITTTTVTPTPTGTQTPTTTPITTTTTVTPTP-TGTQTPTT  2243

Qy     794  TPFKITTLKTTTLAPKVTTT--KKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKA  852
              ||  |   |:|  |          | |||                ||   |   || |
Db     2244 TP--ITT--TTTVTPTPTPTGTQTPTTTPITTTTVTPTPTPTGTQTPPTPITTTTTVT  2299

Qy     853  P-KKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNP--TSRIAEAMLQTTTR-PNQT  907
              |   ||:    | | ||  ||||   || |:|  |:   |  |  |     ||| ||:
Db     2300 PTPTPTGTQTPTTTPITTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTT  2359

Qy     908  PNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTP  945
              |:   |:                ||:
Db     2360 PITTTTTVTPTPTPTG---TQTP--TTTPITTTTVTP  2392
```

```
Qy   881  -------TSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHM- 932
                  ::|                 :||    |||:.:||
Db  1454  YLRGYPDGSFRPERNITRAEAAVIF---------AKLL---GADESYGAQSASPYSD 1498

Qy   933  LLRPH------------------------------VFMPE-------VTPDMDYLPRVPNQGI----- 958
          ||:                              ||:            :|:||:|
Db  1499  LADTHWAAWAIKFATSQGLFKGYPDGTFKPDQNITRAEFATVVLHFLTKVKGQEIMSKLA 1558

Qy   959  ---IINPMLSDETNICNG----KPVDGLTTL----RNGTLVAFRGHYFWMLSPFSPPS 1005
              :|  |      :: |||:||         :||       |||
Db  1559  TIDISNPKFDD-----CVGHWAQEFIEKLTSLGYISGYPDGT-----------FKPQN 1600

Qy  1006  PARRITEVWGIPSPIDTVFTRCNCEGKTFFFKD--SQYWRFTNDIKD 1050
                  |        :: |       ||     |||  |||
Db  1601  YIKRSESV----ALINRALERGPLNGAPKLFPDVNESTWAF-GDIMD 1642


RESULT 3
AMYH_YEAST
ID   AMYH_YEAST      STANDARD;       PRT;   1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
DE   GLUCOSIDASE) (1.4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule S., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
CC       GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
CC       WITH RELEASE OF BETA-D-GLUCOSE.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13C.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; Z38061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
```

```
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA32069.1; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; S0001458; MUC1.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL           1     21       POTENTIAL.
FT   CHAIN           22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN         210    817       SER/THR-RICH.
FT   CARBOHYD       817    817       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD       874    874       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;


Query Match             12.0%;  Score 789.5;  DB 1;  Length 1367;
Best Local Similarity   28.4%;  Pred. No. 7.9e-26;
Matches  301;  Conservative 104;  Mismatches 483;  Indels 171;  Gaps 47;


Qy    59  TTPDSTTQHNKVSTSPKITTAKPINP------------RPSLPPNSDTSKETSLTVNKETTVE 110
          ::|:|:|:|  :|     ||:.:||             :|:.|       |:  ::|:
Db   236  STSESSTTTSSTSBSSTSSSSTTAPATPTTTSCTKEKPTPPTTTSCTKEKPTPPHHDTTPC 295

Qy   111  TKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETT--TKGPA 168
          ||:. |||: | : |    :||| : ||   |          PTPSSSTTESSSAPV
Db   296  TKKKTTTSK-TCT--KTTTPVPTPS-SSTTESSSAPV------PTPSSSTTESSSAPV 344

Qy   169  LTTPKE------PTPTTPKEPASTTPKEPTPTTIKSAP---TTPKEPAPTTTKSAPTTP 218
                       |:|:|:||: |   :  ||   |:|      |: |  ||   |:
Db   345  TSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTPSSSTTE 404

Qy   219  KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKEPAP 278
              ||   |  ||   |  ||   :|| :|||| |    :|| ||  |:|  :|
Db   405  SSTTESSSAPVTSS----TTESSSAPVT----SSTTESSSAPVTSSTTESSSAPVTSSTTES 453

Qy   279  TTPKEPTP---TTPKEPAPTTTKEPAPTTPKEPAPTTPAPKPAPTTPKEPAP-----TTPKEP 331
          :: ||| |   |||: |    :|||   ||   ||
Db   454  SSAPVPTPSSSTTESSSAPVT--SSTTESSSAP-VPTPSSSTTESSSAPVTSSTTESSS 509

Qy   332  APTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAP----TTTKSAPTTPKEPSPTTTKEPAP 387
          || | |  | | ||| ||     :| : ||       |:|||| |      | |:
Db   510  APVPTPS-SSTTESSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSSTP 568

Qy   388  ---TTPKEPAPT-TPKKPAPTTPKEPAPTTTKKPAPTTTKKPAPTAPKEPAPTTPKETA 442
             || ||| |  | ||| |    ||| |  :: |||     :|| ||   || |
Db   569  VTSSTTESSSAPVPTPSSSTTESSSTTESSSAPVPT----PSSSTTESSSAPAPTTPSSSTTESSSA 624

Qy   443  PTTPKKLTPTTPEKLAPTTTPEKPAPTTPEELAPT-TPEEPTPTTPEEPAP------TTPKA 496
          |:|           ||   | ||  |  ||    | |      || ||         || |
Db   625  PVT------SSTTESSSAP-VPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESS 679

Qy   497  AAPNTKPEPAPTTPKEPAPTTPKEPAPTTPKETAPT-TPKGTAPTTLKEPAPTTPKKPAP 555
          :|| |    ||| |    ||   |      ||::||:| |||:|
Db   680  SAPVT----SSTTESSSAPVT---SSTTESSSAPVTPSSSTTESSSAPVTPSSSTTE 731

Qy   556  KELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAP---TTPKEPAPT-TPKGTAPTTL 611
          || ||   | |  | ||: ||   |:||   |        ||  || |  | ||:|
Db   732  SSSAPVPTPS-SSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESS 790

Qy   612  KEPAPTTPKKPAPKELAPTTKGPTSTTSDKPAPTTPKETAPTTPKEPAP-----TTPKK 666
              ||   |:|     ||| |   | :||     ||:||:|
Db   791  SAPVPTPSSSTTESSVAPVPTPSSSSNITSSAPSSTPFSSSTESSSVPFSSSSTTESS 850

Qy   667  PAPTTPETPPPTTSEVSTPT-----TTKEPTTIHKSPDESTPELSAEPTPKALENSPKEP 721
          ||| |     || | :||:       |||:|       : |  |:  |  |::
Db   851  SAPVSSSTTESSSVAPVPTPSSSSNITSSAPSSI---PFSSTSSFSTGT-TVTPSSSKYP 906

Qy   722  GVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITAT 781
          | | :|  ||| || |  |: ||: | | |:|: | |  |  |:  ||  ||: :|
Db   907  GSQTETSVSSTTETTTIVPTKTTTSVTTPSTTTTTTVCSTGTNSAGETTSGCSPKTVTTT 966

Qy   782  --TTQVTSTTQDTTPFKITTLKTTTLAPKVTT---TKKTITTT-EIMNKPEETAKPKDR 835
            |:|:|:||:| |||: ||:|  ||:|||:||     ||:|||| :|::|  ||:| |:
```

```
Db    967  VPTTTTTSVTTSSTTTTTTVCSTGTNSAGETTSGCSPKTITTTVPCSTSPSETA-----  1021
                || ||  ||    : |||    | |  ||   :       ||   | |||||
Qy    836  ATNSKATTPKPQKP-------------------TKAPKKPTSTKKPKTMPRVRKPKTTPTPR  878
               |:|||  |       | |       : :|     :|:|    | | ||
Db   1022  ---SESTTTSPTTSPVTTVVSTTVVTTEYSTSTKPGGITTTFVTKNIPTTYLTTIAPTP-  1077
              :  :||       ||:  | | ||  |  | |   :||   | | ||
Qy    879  KMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHV   938
              :   :|||    | |       |        ||   :     ||
Db   1078  -SVTTVTNFTPTT-ITTTVCSTGT-----------------NSAGETTSGCSPKTVTTTVPCST  1122
               |        || |                   |:  |||
Qy    939  FMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTTLRNGTLVAFRGHYFWML   998
                |   :                            ||    |   ||||
Db   1123  GTGEYTTEATTLVTTAVTTTVVTTESSTGTNSA-GKTTTGVTTKSVPTT------YVTTL   1175
              :|:  ||                          || |    ||    |:|
Qy    999  SPFSPPSPARRITEWGIPSPIDTVFTRC----NCEGKT  1033
              :|:  :||                :|:  || |   |:|
Db   1176  APSAPVTPATN-----AVPTTITT--TECSAATNAAGET  1207
```

RESULT   4
MUC1_HUMAN
ID   MUC1_HUMAN   STANDARD;   PRT;   1255 AA.
AC   P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM)
DE   (EPISIALIN) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE   (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-
DE   REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE   DF3).
GN   MUC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=90368716; PubMed=2394722;
RA   Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT   "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL   J. Biol. Chem. 265:15294-15299(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90202794; PubMed=2318825;
RA   Ligtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT   "Episialin, a carcinoma-associated mucin, is generated by a
RT   polymorphic gene encoding splice variants with alternative amino
RT   termini.";
RL   J. Biol. Chem. 265:5573-5578(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90368715; PubMed=1697589;
RA   Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA   Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.,
RT   "Molecular cloning and expression of the human polymorphic epithelial
RT   mucin.";
RL   J. Biol. Chem. 265:15286-15293(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=91097524; PubMed=2268309;
RA   Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA   Taylor-Papadimitriou J., Gendler S.J.;
RT   "Structure and expression of the human polymorphic epithelial mucin
RT   gene: an expressed VNTR unit.";
RL   Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90276413; PubMed=2351132;
RA   Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
RA   Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA   Keydar I.;
RT   "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT   may generate multiple protein forms.";
RL   Eur. J. Biochem. 189:463-473(1990).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90276414; PubMed=2112460;
RA   Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA   Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT   "A transcribed gene, containing a variable number of tandem repeats,
RT   codes for a human epithelial tumor antigen. cDNA cloning, expression
RT   of the transfected gene and over-expression in breast cancer
RT   tissue.";
RL   Eur. J. Biochem. 189:475-486(1990).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91033045; PubMed=1688329;
RA   Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA   Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT   "Isolation and characterization of an expressed hypervariable gene
RT   coding for a breast-cancer-associated antigen.";
RL   Gene 93:313-318(1990).
RN   [8]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=88330762; PubMed=3417635;
RA   Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA   Burchell J.;
RT   "A highly immunogenic region of a human polymorphic epithelial mucin
RT   expressed by carcinomas is made up of tandem repeats.";
RL   J. Biol. Chem. 263:12820-12823(1988).
RN   [9]
RP   SEQUENCE OF 1-169 FROM N.A.
RX   MEDLINE=90088473; PubMed=2597151;
RA   Abe M., Siddiqui J., Kufe D.;
RT   "Sequence analysis of the 5' region of the human DF3 breast
RT   carcinoma-associated antigen gene.";
RL   Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN   [10]
RP   SEQUENCE OF 1-109 FROM N.A.
RC   TISSUE=Thyroid;
RX   MEDLINE=96183746; PubMed=8608966;
RA   Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT   "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT   transcriptase polymerase chain reaction of the MUC1 gene.";
RL   Int. J. Cancer 66:55-59(1996).
RN   [11]
RP   SEQUENCE OF 1-89 FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=96181716; PubMed=8604237;
RA   Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA   Lee L.N., Luh K.T., Wu C.W.;
RT   "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT   tissues.";
RL   Oncology 53:118-126(1996).
RN   [12]
RP   SEQUENCE OF 1-46 FROM N.A.
RC   TISSUE=Breast carcinoma;
RA   Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
RL   Submitted (OCT-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC   CYTOSKELETON.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
CC   IS ALSO PRODUCED.
CC   -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
CC   SPLICING.
CC   -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
CC   TUMORS, SUCH AS BREAST CANCER.
CC   -!- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND SIALIC.
CC   ACID).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT
```

```
Qy    158  P-KAETTTKGP-ALTTPKEP------------------------TPTTPKEPASTTPKEPTPTTI-  197
Db     66  PGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTPALGSTTPPAHDVTSAPDNK-       125

Qy    198  SAPTTPKEPAPTTKSAPTTPKEPAPTTKEP----------APTTPKEPAPTTTKEPAPTTTK-    251
Db    126  PAPGSTAPPAHGVT-SAPDT--RPAPGSTAPPAHGVT-                            179

Qy    252  SAPTTPKEPAPTTP-----------KKPAPTTPKEP-------APTTPKEPTPTTP-          289
Db    180  SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT      239

Qy    290  ----KEPAPTTKEP--------APTTPKEPAPTTPK---KPAPTTPKEPAPTTPK------E    330
Db    240  SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT-    299

Qy    331  PAPTTTKEPSPTTPKEPAPTT--TKSAPTTPKEPAPTTTTKSAPTTPKEPSPTTTKEP-      385
Db    300  SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT-SAPDT--RPAPGSTAPPAH    356

Qy    386  ----APTTPKEPAPTTPK---KPAPTTPKEPAPTTPK----EPAPTTTKKPAPTAPKEP     433
Db    357  GVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAH    416

Qy    434  APTTPKETAPTTPKKLTPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTT-           493
Db    417  GVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP----PAHGVTSAPDTRPAPGSTA    472

Qy    494  PKA---AAPNTPKEPAPTTPK---EPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP    545
Db    473  PPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP----PAHGVTSAPDTRPAP    528

Qy    546  APTTP----------KKPAPKELAPTTKEPTSTTSDKPAP--TTPKGTAPTTPKEPAPT    593
Db    529  GSTAPPAHGVTSAPDTRPAPGSTAP-PAHGVTSAPDTRPAPGS-TAPPAHGVTS----APD    583

Qy    594  TPKEPAPTTPKG---TAPTTLKEPAPTTP------------KKPAPKELAPTTKGPTSTT    639
Db    584  TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP-PAHGVTSAP    642

Qy    640  SDKPAP--TTPK---ETAPTTPKEPAPTTP------------KKPAPTTPETPPPTTSEVS    683
Db    643  DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGS--TAPPAHGVTS    700

Qy    684  TPTTTKEPTT--------IHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEM    736
Db    701  APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP----DTRPAPGSTAP--    753

Qy    737  TTTAKDKTTERDLRTTPETT------TAAPKMTKETATTTEKTTESKITATTTQVTSTT    789
Db    754  --PAHGVTSAPDTRPAPGSTAPPAHGVTSAP---DTRPAPGSTAPPAHGVTSAPDTRPA    807

Qy    790  TQDTTP-FKITTLKTTLAPKVTTTKKTITTT--EIMNKPEETAKPKDRATNSKATTPK    845
Db    808  PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA    867

Qy    846  PQKPTKAPKKPTSTKKPKTMPRVRKPKTTPTPPRKMTSTMPELNPTSRIAEAMLQTTTRPN    905
Db    868  P--GSTAPPAHGVTSAPDTRP--APGSTAPPAHGVTSAPDTRPA-----------PG    909

Qy    906  QT-PNSKLVEVNPKSEDAGAGEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGILINPML    964
Db    910  STAPPAHGVTSAPDTRPAPGSTAPPAH-----------GVTSAPDNRPALGSTA----PPV    955

Qy    965  SDETNICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPS 1005
Db    956  HNVTSASGSASGSASTLVHNGTSARATTTPASKSTPFSIPS 996
```

RESULT   5
FP1_MYTED
ID   FP1_MYTED   STANDARD;   PRT;   875 AA.

CC   VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
CC   MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; J05582; AAA60019.1; -.
DR   EMBL; M32738; AAA35804.1; -.
DR   EMBL; M32739; AAA35806.1; -.
DR   EMBL; J05581; AAA59876.1; -.
DR   EMBL; M61170; AAB53150.1; -.
DR   EMBL; X52229; CAA36478.1; ALT_SEQ.
DR   EMBL; X52228; CAA36477.1; ALT_SEQ.
DR   EMBL; M35093; AAB59612.1; ALT_SEQ.
DR   EMBL; Z17324; CAA78972.1; -.
DR   EMBL; Z17325; CAA78973.1; -.
DR   EMBL; M31823; AAA35757.1; -.
DR   EMBL; S81781; AAD14376.1; ALT_INIT.
DR   EMBL; S81736; AAD14369.1; ALT_INIT.
DR   EMBL; M21868; AAA59874.1; ALT_SEQ.
DR   PIR; A35175; A35175.
DR   PIR; B35175; B35175.
DR   PIR; S10218; S10218.
DR   GlycoSuiteDB; P15941; -.
DR   MIM; 158340; -.
DR   MIM; 113720; -.
DR   InterPro; IPR000082; SEA.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW   Repeat; Alternative splicing.
FT   SIGNAL   1   23
FT   CHAIN   24   1255
FT   DOMAIN   24   1162   POTENTIAL.
FT   TRANSMEM   1163   1186   MUCIN 1.
FT   DOMAIN   1187   1255   EXTRACELLULAR (POTENTIAL).
FT   DOMAIN   81   960   POTENTIAL.
FT   DOMAIN   1034   1151   CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD   957   957   44 X 20 AA TANDEM REPEATS.
FT   CARBOHYD   975   975   SEA.
FT   CARBOHYD   1029   1029   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1055   1055   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1133   1133   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   19   19   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   20   22   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   20   31   T -> TATTAPKPAT (IN ISOFORM B).
FT   VARSPLIC   126   905   MISSING (IN ISOFORM C).
FT   MISSING (IN ISOFORM D).
FT   MISSING (IN POLYMORPHIC EPITHELIAL
FT   ISOFORM).
FT   VARSPLIC   1077   1087   FLQIYKQGFL -> VSIGLSFPMLP (IN SECRETED
FT   ISOFORM).
FT   VARSPLIC   1088   1255   MISSING (IN SECRETED ISOFORM).
FT   CONFLICT   2   2   T -> A (IN REF. 11).
FT   CONFLICT   134   134   P -> Q (IN REF. 9).
FT   CONFLICT   154   154   P -> Q (IN REF. 9).
FT   CONFLICT   1021   1021   S -> T (IN REF. 3).
FT   CONFLICT   1251   1251   A -> T (IN REF. 3).
SQ   SEQUENCE   1255 AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;

Query Match   9.9%; Score 651; DB 1; Length 1255;
Best Local Similarity 27.8%; Pred. No. 3.3e-20;
Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;

Qy   99 TSLTVNKETTVETKETTTTNKQTSDGKEKTTSAKETQSIEKTSAKD-LAPTSKVLAKPT 157
                 |||| : |||||| : || |: :|: |::| |||::
Db   16 TVLTV--------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS   65

```
AC   Q25460;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
DE   PROTEIN 1) (MEFP1) (FRAGMENT).
GN   FP1.
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=6550;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91025829; PubMed=1367451;
RA   Filpula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
RT   "Structural and functional repetition in a marine mussel adhesive
RT   protein.";
RL   Biotechnol. Prog. 6:171-177(1990).
RN   [2]
RP   PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
RX   MEDLINE=83157732; PubMed=6298211;
RA   Waite J.H.;
RT   "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
RT   hydroxyproline-containing decapeptide in the adhesive protein of the
RT   mussel, Mytilus edulis L.";
RL   J. Biol. Chem. 258:2911-2915(1983).
CC   -!- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC       PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC       ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC       FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC   -!- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC   -!- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-K IS POST-TRANSLATIONALLY
CC       MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
CC       HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
CC       (DOPA) DERIVED FROM TYROSINE.
```

```
DR   EMBL; X54422; CAA38294.1; -.
DR   InterPro; IPR002964; Adhesive_plaq.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01216; ADHESIVEI.
DR   PRINTS; PR01217; PRICHEXTENSN.
KW   Repeat; Hydroxylation.
FT   NON_TER       1      1
FT   DOMAIN       67    870       TANDEM REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
FT                                P-P-[ST].
SQ   SEQUENCE   875 AA;  100412 MW;  6EA85312748CAACE CRC64;

Query Match             8.5%;    Score  556.5;  DB 1;  Length 875;
Best Local Similarity  28.7%;    Pred. No. 1.7e-16;
  Matches  272;  Conservative 106;  Mismatches 375;  Indels 195;  Gaps 56;

Qy    66 TQHNKVSTSPKITTAKPINP----------------RP--SLPPNSDTSKETSLITVNKETTVET 111
Db     1 TKHEPVYKPVKTSYSAPYKPPTYQPLKKKVDYRPTKSYPPTYG--SKTNYLPLAKKLSSYK 59

Qy   112 KETTTNKQTSTDG--KEKTT---SAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKG 166
Db    60 PIKTTYNAKTNYPPVYKPKMTYPPTYKPKSYPPTYKPKSYPPTYKPKITYPPTYKAKPSY-- 117

Qy   167 PALTTPKEPTPTTPKE----PASTTPKEPTPTTIKSAPTTP-----KEPAPTTTKSAPTT 217
Db   118 PSSYKPKTYPPTYKPKLTYPPTYKPKPKSYPPTYKPKSYPPSYKTKKTYPSSYKAKPSY 177

Qy   218 PKEPAPTTTKEPA-------PTTTKSAPTTPK--EPAPTT 264
Db   178 P----PTYKAKPSYPPTYKAKPTYSTYKAKPSYPPTYKAKPTY 233

Qy   265 PKKPA--PTTPKEPA--PTTPKEPT--PTTPKEP---APTTKEPAPTTPKEPA--PTAPK 313
Db   234 KAKPSYPPTYKAKPSYPPTYKAKPTYKAKPTYKAKPTYKAKPSYPPTYKAKPSYPPTYKA 293

Qy   314 KPA--PTTPKEPA--PTTPKEPA--PTTKEPS--PTTPKEPA--PTTTKSAPTTTKEPA- 363
Db   294 KPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKPTY 353

Qy   364 PTTTKSAPTTPKEPSPTTTKEPAPTTPKKPA--PTTPKEPA--PTTPKEPAP 419
Db   354 PSTYKAKPSYP----PTYKAKPSYPPTYKAKPTYKAKPSYPPTYKAKPSY 409

Qy   420 TTTKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEELAPTTPE 479
Db   410 PPTYKAKPTYKAKP-------TYPST-YKAKPSYP---PSYKAKPSYPPTYKAKPTYKA 457

Qy   480 EPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKE----PAPTTPKETAPTT-- 533
Db   458 KPT-------YPSTYKA----KPSYPASYKAKPSYPPTYKSKSSYPSSYKPKKTYPPTYK 506

Qy   534 PKGTAPTTLKEPAPTTPKKPAPKELAPTTKEPTSTTSDKPAPTTPKGTAPTT---PKE 589
Db   507 PKLTYKPTYK-PKPSYPPSYKPKTYPPTYK------------PKISYPPTYKAKPSY 551

Qy   590 PAPTTPKEPAPTTPKGT--APTTLKEPA--PTTPKKPA--PKELAPTTTKG-PTSTTSD 641
Db   552 PATYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYPSTYKAKPSYPPTY 611

Qy   642 KPAPTTPKETAPTTPKEPA--PTTPKKPA-PTT-----PETPPPTTSEVSTPTTTKE 690
Db   612 KAKPSYP----PTYKAKPSYPPTYKAKPTYPSTYKAKPSYPPTYKPKISYPPTYKAKPSY 667

Qy   691 PTTIHKSPDESTPELSAEPTPKALENSPK----EPGVPTT--KTPA----ATKPEMTTT 739
Db   668 PPT-YKAKPSYPPTYKAKPTNPSTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPPT 726

Qy   740 AKDKTTERDLRTTPETTTAAPKMTKETATTEKTESKITATTTQVTSTTQDTTPFKIT 799
Db   727 YKAKPTYKAKPTYPSTYKAKP--TYKAKPTYPPTYKAK-------PSYPPTYKPKPSYPP 777

Qy   800 TLKTTTLAPKVITTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTRAPKK---P 856
Db   778 TYKSKSIYPSSYKPKKTYPPT--YKPKLTYPPTYK---------PKPSYPPSYKPKITYP 826

Qy   857 TSTKKPKTMPRVRKPKTTPTP----------------RKMTSTMPELNPTSR 892
Db   827 STYKLKPSYPPTYKSKTSYPPTYNKKISYPSSYKAKTSYPPAYKPTNR 874


RESULT 6
ZAN_HUMAN
ID   ZAN_HUMAN      STANDARD;      PRT;  2700 AA.
AC   Q9Y493; O00218;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ZONADHESIN (FRAGMENT).
GN   ZAN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-2379 FROM N.A.
RX   MEDLINE=99018118; PubMed=9799793;
RA   Glockner G., Scherer S., Schattevoy R., Boright A., Weber J.,
RA   Tsui L.C., Rosenthal A.;
RT   "Large-scale sequencing of two regions in human chromosome 7q22:
```

```
RT       analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT       reveals 17 genes.";
RL       Genome Res. 8:1060-1073(1998).
RN       [2]
RP       SEQUENCE OF 2338-2700 FROM N.A.
RC       TISSUE=Testis;
RX       MEDLINE=97271566; PubMed=9126492;
RA       Gao Z., Harumi T., Garbers D.L.;
RT       "Chromosome localization of the mouse zonadhesin gene and the human
RT       zonadhesin gene (ZAN).";
RL       Genomics 41:119-122(1997).
CC   -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC       OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC       SIGNALING.
CC   -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC       APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC   -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC       ZONA PELLUCIDA.
CC   -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC       THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC       SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC   -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC       OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC   -!- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC       ----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute.  There are no restrictions on its
CC       use  by  non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       ----------------------------------------------------------------
DR   EMBL; AF053356; AAC78790.1; -.
DR   EMBL; U83191; AAC51208.1; -.
DR   MIM; 602372; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000998; MAM.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR002919; TIL.
DR   InterPro; IPR003328; TILa.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00629; MAM; 3.
DR   Pfam; PF01826; TIL; 4.
DR   Pfam; PF02345; TILa; 4.
DR   Pfam; PF00094; vwd; 4.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00137; MAM; 2.
DR   SMART; SM00214; VWC; 1.
DR   SMART; SM00011; VWC_def; 3.
DR   SMART; SM00216; VWD; 3.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS50060; MAM_2; 4.
KW   Glycoprotein; Transmembrane; Cell adhesion; Repeat.
FT   NON_TER       1       1
FT   DOMAIN        1     109       MAM 1.
FT   DOMAIN      <1     112       MAM 2 (PARTIAL).
FT   DOMAIN     112     136       MAM 3.
FT   DOMAIN     161     326       MAM 4.
FT   DOMAIN     322     446       66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT   DOMAIN     483     951       (MUCIN-LIKE DOMAIN).
FT   DOMAIN     953    1065       VWFD 1 (PARTIAL).
FT   DOMAIN    1066    1454       VWFD 2.
FT   DOMAIN    1455    1861       VWFD 3.
FT   DOMAIN    1862    2292       VWFD 4.
FT   DOMAIN    2293    2684       VWFD 5.
FT   DOMAIN       ?       ?       EGF-LIKE.
FT   CARBOHYD    74      74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   403     403       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1023    1023       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1099    1099       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1618    1618       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1737    1737       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1832    1832       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1878    1878       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  2136    2136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  2505    2505       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT  2374    2379       NNQKMA -> RAGPGP (IN REF. 1).
FT   NON_TER   2700    2700
SQ   SEQUENCE OF 2700 AA;   293013 MW;  80E60CC0B12277B1 CRC64;
```

```
Query Match                           8.5%;     Score 555.5; DB 1;    Length 2700;
Best Local Similarity   32.6%;   Pred. No. 4.9e-16;
Matches  199;   Conservative  64;   Mismatches 240;   Indels 107;   Gaps  32;

Qy      82  PINPRPSLPPNSDTSKETSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKT  141
                : ||||||  ||   ||  :  ||| |||  |      ||| |  ||:|      ||
Db     445  PVKVLPELPPVSPVSS---TGPSETTGLTENPTISTK--------------KPTVSIEKP  487

Qy     142  SAKDLAPTSKVLAKPTPKAETTKGPALTTPKEPTPTPKEPASTTPKEPT-----PTTI  196
                    |   |  |   :|| ||||:|| || :|  ||  ||   |     |||
Db     488  SVTTEKPT-----VPKEKPTIPTEKPTISTEK--PTIPSEKPNMPSEKPTIPSEKPTIL  539

Qy     197  KSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTPKEPA-PTTTKSAPT  255
                || ||  ||  ||   || |   ||  |  |:| | :||  || ||
Db     540  TEKPTIPSE-KPTIPSEKPTISEKPTVPTEE--PTTPTEETTYMEEPVIPTEKPSIPT  596

Qy     256  TPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTPKEPAPTTKEPAPTTKEPAPTTAPKKP  315
                :  |  |   || | :| | ||   | ||| :  ||  |   || |: ||
Db     597  --EKPSIPTEK--PTISMEETIISTEKPTICPEKPTIPTEK---PTIPEKSTISPEK-  647

Qy     316  APTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTKEPAPTTTKSAPTTPK  375
                |||  | |  :| ||   |  :  | |      || |    || ||   || ||
Db     648  -PTTPTE-KPTIPTEKPTISTEKPTIPTEK-PTISPEKLTIPTEKLTIPT---EKPTIPT  701

Qy     376  EPSPTTTKEPAPTTPKEPAPTTPKKPA------PTTPKEPAPTTPKEPAPTTKKPPAPTAP  430
                | :|| | :| | | |  :|   |    |    PTTPKEPAPTTPKEPAPTTKKPAPTAP
Db     702  EKPTISTEE--PTTPTEETTISTEKPSIPMEKPLPTEETTISVEETTISTEKLTIPM-  757

Qy     431  KEPAPTTPKETAPT------TPKKLIPTTPEKLA------PTTPEKPAPTTPEELAPTTPEE  480
                | ||||       | |   || |   |   ||  |      PTTPEKPAPTTPEELAPTTPEE
Db     758  EKPTISTEKPTIPTEKPTISPEKLTIPT-EKLTIPTEKPTIPIEETTISTEKL--TIPTE  814

Qy     481  PTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPT  540
                  |  || :|  ||: |  ::| || |  ||  ||||: || ||  |||  |
Db     815  KPTISPEKPTISTEK-----------PTIPTE-KPTIPTE-------ETTISTEKLTIPT  855

Qy     541  TLKEPAPTTPKKPAPKELAPTTTKEPTSTT------SDKPAPTTPKGTAPTTPKEPAPTT  594
                ::| || | || | ||: | | |:|| |  ||  :  || |  | ||| || ||
Db     856  --EKPTISPEKLTIPTEKPTISTEKPTIPTEKPTIPT-EKLTALR  911

Qy     595  PKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKGPTSTTSDKPAPTTPKET--A  652
                  |  || | |   || || || | || || |  ||  | |   || ||  ||  A
Db     912  PPHPSPTA-TGLAALVMSPHAPSTPMTSV--ILGTTTTSRSSTGMSCP-PNARYESCAC  966

Qy     653  PTTPKEPAPT  662
                 |  | |:
Db     967  PASCKSPRPS  976
```

```
RESULT    7
EXTN_TOBAC
ID   EXTN_TOBAC   STANDARD;   PRT;   620 AA.
AC   P13983;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
GN   HRGPNT3.
OS   Nicotiana tabacum (Common tobacco).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. XANTHI; TISSUE=Leaf;
RX   MEDLINE=90128263; PubMed=2612909;
RA   Keller B., Lamb C.J.;
RT   "Specific expression of a novel cell wall hydroxyproline-rich
RT   glycoprotein gene in lateral root initiation.";
RL   Genes Dev. 3:1639-1646(1989).
CC   -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC       THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC       MAIN ROOT.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC   -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC       SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC       GLYCOSYLATED.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X13885; CAA32090.1; -.
DR   PIR; S06733; S06733.
KW   Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW   Hydroxylation.
FT   SIGNAL        1      ?
FT   CHAIN         ?    620
FT   REPEAT       70     73       EXTENSIN.
FT   REPEAT      148    151       H-A-P-P.
FT   DOMAIN      229    242       H-A-P-P.
FT   REPEAT      229    235       2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT   REPEAT      236    242       1.
FT   DOMAIN      205    620       2.
FT   DOMAIN      499    600       CONTAINS THE SER-PRO(4) REPEATS.
FT                                3 X APPROXIMATE TANDEM REPEATS.
SQ   SEQUENCE   620 AA;  65406 MW;  641DD2278AB28524 CRC64;
```

```
Query Match          8.4%;  Score 551; DB 1; Length 620;
Best Local Similarity  27.2%; Pred. No. 2.1e-16;
Matches  170;  Conservative  64;  Mismatches  304;  Indels  88;  Gaps  24;

Qy  160  AETTTK-GPALTTP--KEPTPTT----PKEPASTTPKE---PTPTTIKSAPTTPKEPAP 208
Db   24  AEATTQYGGVLPPPVTSQPPPSSIGLSPPSAPTTTPPSRGHVPSP---RHAPPRHAYPPP  80

Qy  209  TTTKSAPTTPKEP------APTTTKEPAPT---TPKEPAPTTTKEPAPTTTKSAPTTPKE 259
Db   81  SHGHLPPSVGGPPPHRGHLPPSRGFNPPPSPVISPSHPPSIYGAPPSHGPGHLPSHGQR 140

Qy  260  PAPTTPKKPAPT---TPKEPAPTTPKEPTPTTPKEPAPTTK--EPAPTTPKEPAPTAPKK 314
Db  141  P------PSPSNGHAPPSGGHPPRGGHPPSHRRPSPPSRGHPPPSYSPPPPSIYS      193

Qy  315  PAPTTPKEPAPTTPKEPAPTTTKEPSPTTPK---EPAPTTKSAPTTTKEPATTTKS-- 369
Db  194  PSPQV--QPPT--YSPPPTHVQPTPSPSRGHQPQPPTHRHAPPTHRHAPPTHQPSPL 249

Qy  370  ---APTTPKEPSPTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPT---TPKEPAPTTK 423
Db  250  RHLPPSPRRQPQPPTYSPPPAVAQSPQSPPYSPPPPTYSPPPPSIYSPPPPAYSPSP 309

Qy  424  KPAPT---APKEPAPTTPKETAPTTPKKL-TPTT-----PEKLAPTTPEKPAPTTPEEL 473
Db  310  PPTPTPTFSPPPPAYSPPPTYSPPPPTYLPLPSSPIYSPPPPVYSPPPPVYSPPPPTYL 369

Qy  474  APTTPEEPTPTTPEEPAPTTPKAAAP---NTPKEPAPTTPKEPAPTTPKEPAPTTPKETA 530
```

```
Db  370  PPPPSSPPPFSPPPPTYEQSPPPPAYSPPLPAPPTYSPPPT--YSPPPPTYAQPP  427
Qy  531  PTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKEPTSTTSDKPAPTTPKGTAPTTPKEP 590
Db  428  P------LPPTYSPPPPAYSPPPPTYSPPPPTYSPPPPAVAQPPPPPTYSPPPPAYSP 481
Qy  591  APTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKGPTSTTSDKPAPTTPKE 650
Db  482  PPPSPIYSPPPPQVQPLPPTFSPPPPRRIHLPPPHRQP---RPPTPTYGQPPSPPTFSP 538
Qy  651  TAPTTPKEPAPT--TPKKPAPTTPETPPPTTSEVSPTTTKEPTTIHKSPDESTPELSAE 708
Db  539  PPPRQIHSPPPHWQPRTPTTYGQPSPPTFSAPPPRQIHSPPPPHRQPRPPTTYGQP    598
Qy  709  PTPKALENSPKEPGVPTTKTPAATKP 734
Db  599  PSP--------PTTYSPPSPPP  612
```

```
RESULT   8
NFH_MOUSE
ID   NFH_MOUSE     STANDARD;     PRT;   1087 AA.
AC   P19246; Q61959;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE   (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN   NEFH OR NFH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89121513; PubMed=3220257;
RA   Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
RA   Mushynski W.;
RT   "Sequence and structure of the mouse gene coding for the largest
RT   neurofilament subunit.";
RL   Gene 68:307-314(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89089138; PubMed=3145094;
RA   Shneidman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
RT   "The structure of the largest murine neurofilament protein (NF-H) as
RT   revealed by cDNA and genomic sequences.";
RL   Brain Res. 464:217-231(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN-SWISS WEBSTER; TISSUE=Brain;
RA   Carden M.J.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC       NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC       SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC       TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; M24496; AAA39813.1; JOINED.
DR   EMBL; M23349; AAA39813.1; JOINED.
DR   EMBL; M24494; AAA39813.1; JOINED.
DR   EMBL; M24495; AAA39813.1; JOINED.
DR   EMBL; M35131; AAA39809.1; ALT_FRAME.
DR   EMBL; Z31012; CAA83229.1; -.
DR   PIR; JT0368; QFMSH.
DR   PIR; A43778; A43778.
DR   MGD; MGI:97309; Nfh.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Repeat.
FT   DOMAIN        1     97       HEAD.
FT   DOMAIN       98    408       ROD.
FT   DOMAIN      409   1087       TAIL.
FT   DOMAIN      436    517       GLU-RICH (ACIDIC).
FT   DOMAIN      519    886       50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
FT   DOMAIN      887   1087       GLU/LYS-RICH.
FT   DOMAIN       98    129       COIL 1A.
FT   DOMAIN      130    141       LINKER 1.
FT   DOMAIN      142    239       COIL 1B.
FT   DOMAIN      240    261       LINKER 12.
FT   DOMAIN      262    283       COIL 2A.
FT   DOMAIN      284    287       LINKER 2.
FT   DOMAIN      288    408       COIL 2B.
FT   CONFLICT    133    133       K -> QA (IN REF. 2 AND 3).
FT   CONFLICT    199    199       A -> AR (IN REF. 2 AND 3).
FT   CONFLICT    281    281       S -> T (IN REF. 2 AND 3).
FT   CONFLICT    492    492       L -> G (IN REF. 2 AND 3).
FT   CONFLICT    551    551       P -> PREAKSP (IN REF. 3).
FT   CONFLICT    689    712       MISSING (IN REF. 3).
FT   CONFLICT    714    714       G -> A (IN REF. 3).
FT   CONFLICT    814    814       V -> M (IN REF. 2 AND 3).
FT   CONFLICT    843    843       T -> N (IN REF. 2 AND 3).
SQ   SEQUENCE   1087 AA;  116612 MW;  57BAC76A38ED1CB9 CRC64;

Query Match          8.1%;
Best Local Similarity  27.5%;    Score 533;   Pred. No. 1.9e-15;   DB 1;   Length 1087;
Matches  216;   Conservative  81;   Mismatches 330;   Indels 158;   Gaps 38;

Qy    45  DEAGSGLDNGDFKVT------TPDTSTTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKET   99
          :   : : : :         :   :    :  : :  :   :        :  :    ||
Db   404  EECRIGFGPSFSLTEGLPKIPSIST--HIKVKSEEMIKVVE----------KSEKET    449

Qy   100  SLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPK   159
          :    :   :  : : :        : :   :   : : :   :  :     :    |:
Db   450  VIVEGQTEEIRVTEGVTEEDKEAQGQEGEEAEEGEEKEEELAAATSPPAEEAASPEKE   509

Qy   160  AETTTKGPALTTPKEPTPTTPKEPASTTPKEPPTTTKSAPTTPKEPAPTTTKSAPTTPK   219
          :   :    |  | : | |     : |     : | | | | | | | | | | |
Db   510  TKSRVKEEAKSPGEAKSPA---EAKSPGEAKS-PGEAKSPGEAKSPAEPKSPA       564

Qy   220  EP-APTTKEPAPTTKEPAPTTKEP----APTTKSAPTTKEPA---PTTPKKPA-       269
          : | | | | | | | |      : | |     | | | | | | |
Db   565  EPKSPAEAKSPA--EPKSPA-TVKSPGEAKSPSEAKS-PAEAKSPAEAKSPAE       619

Qy   270  ---PTTPKEPA----PTTPKEP----TPTTPKEPAPTTKEPAPTTPKEPA----PTAPKK   314
          :  : | | |        | | | | |     : |     | | | | | | |
Db   620  AKSPAEAKSPAEAKSPATVKSPGEAKSPGEAKSPA---EAKSPAEAKSPAEVKS       676

Qy   315  P----APTTKEPAPTTPKEPAPTTTKEPSPTTPKEPA----PTTTKSAPTTTKEPAPTT   366
          :  |        | | | | | | | |     : | | | | | | | | | |
Db   677  PGEAKSPAEPKSPAEA--KSPA----EVKSPAEAKSPAEVKSPGEAKS-PAAVKSPAEAK  729

Qy   367  TKSAPTTPKE-PSPTTTKEPA----PTTKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTT   421
          :  : | | | : | |        | | |     : | | |     | |     : |
Db   730  SPAAVKSPGEAKSPGEAKSPAEAKSPAEAKSPAEAKTPVKEGAKSPAEAKSPEK      789

Qy   422  TKKPAPTTAPKEPAPTTPKETAPTTPKKLLTPTTPKELAPTTPKEPAPTTPELAPTTPEEP  481
          | | | |     : | | | |     : | |     : : | | |  : |
Db   790  AKSPVKEDIKPPAEAKSPEKA-KSPVKEGAKPPEKAKPLDVKSPEAQTPVQEEATVPTDI   848

Qy   482  TPTTPEE-PAPTTKAAAPNTPKEPAPTT------PKEPAPTTP------KEPAPTTPKET  529
          :  | : : | |     : | | |     : | | | |
Db   849  RP--PEQVKSPAKEKAKSPE--KEEAKTSEKVAPKKEEVKSPVKEEVKAKEPPKKVEEEK  904

Qy   530  APTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSD-KPAPTTPKGTAPTTP  587
          | | | : : : |     : | |     : : | | : | : | | :
Db   905  TLPTPKTEAKESKKDEAPKEAPKPKVEEKKETPTEKPDSTAEAKKEEAGEKKAVASEE   964

Qy   588  KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTTKGPTSTTSDKPAPTT               647
          :  | | | | | | |     : | | | | | | | | | | | |
Db   965  ETPAKLGVKEEA--KPKEKTETTKTEAEDTKAKEPS---KPTETEKP--------     1006

Qy   648  PKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSA  707
          :  : | | |     : | |
Db   1007 -------------KKEEMPAAPEK-------------KDTKEKTTTESRKPEEK------  1034

Qy   708  EPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMIKETA   767
          | |     : | |     : | |     : | |
Db   1035 ---PK-MEAKVKEDDKSLSKEP--SKPKTEAEKSSSTDQKESOPPE------KTTEDKA  1082

Qy   768  TTTEK  772
          | | | | |
Db   1083 TKGEK  1087

RESULT   9
TCNA_TRYCR
ID   TCNA_TRYCR     STANDARD;      PRT;   1162 AA.
AC   P23253;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SIALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).
GN   TCNA.
OS   Trypanosoma cruzi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5693;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SILVIO X-10/4;
RX   MEDLINE=91277609; PubMed=1711561;
RA   Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
RA   Prioli R.P.;
RT   "The Trypanosoma cruzi neuraminidase contains sequences similar to
RT   bacterial neuraminidases, YWTD repeats of the low density lipoprotein
RT   receptor, and type III modules of fibronectin.";
RL   J. Exp. Med. 174:179-191(1991).
RN   [2]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=91376547; PubMed=1896773;
RA   Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;
RT   "Trypanosoma cruzi: localization of neuraminidase on the surface of
RT   trypomastigotes.";
RL   Trop. Med. Parasitol. 42:146-150(1991).
CC   -!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
CC       PARASITE INVASION OF CELLS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
CC       LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLNEURAMINYL
CC       RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED
CC       NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
CC       GLYCOLIPIDS OR COLOMINIC ACID.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POSSIBLE).
CC   -!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM
CC       IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
```

```
CC   -!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
CC       DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
CC       PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
CC   -!- SIMILARITY: CONTAINS 3 BNR REPEATS.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; M61732; AAA30255.1; -.
DR   PIR; JH0557; JH0557.
DR   HSSP; P29768; 1DII.
DR   InterPro; IPR002860; BNR.
DR   Pfam; PF02012; BNR; 2.
KW   Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
KW   Phosphorylation.
FT   DOMAIN        1    457       CYS-RICH.
FT   REPEAT       23     34       BNR 1.
FT   REPEAT      163    174       BNR 2.
FT   REPEAT      209    220       BNR 3.
FT   DOMAIN      458    588       FIBRONECTIN TYPE-III.
FT   DOMAIN      589   1120       44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT   CARBOHYD    342    342       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    394    394       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1125   1125       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1162 AA;  120032 MW;  07049221897C6A40 CRC64;

Query Match                8.1%;
Best Local Similarity     29.4%;   Score 530.5;  DB 1;  Length 1162;
Matches   177;  Conservative 100;  Mismatches 212;  Indels 113;  Gaps 34;

Qy   176  TPTTPKE-PASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTP  234
          ||:||   :  ::||  || : :||::||  |:  :|| ||  : |||| :||  : :||
Db   599  TPSTPADSSAHSTPSTPVDS--SAHSTPADSSAHGTPSTPVDSSAHGT----PSTP  651

Qy   235  KE-PAPTTTKEPAPTTTKSAPTTPKE------PAPTTPKKP------APTTPKEPA----PTTP  281
          |  |   ||:|  | |:|  ||  |       | :|| |       | ||:|||      :||
Db   652  ADSSAHGTPSTPVDSSAHSTPSTPVDSSAHGAPSTPADSSAHGTPSTP  711

Qy   282  KEPTP-TTPKEPAPTTKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTTKKEPS  340
          ||  | ||:|  | ::|   ||::|  ||  || ::|  |   ||: |::||
Db   712  VDSSAHGTPSTPADSSAHSTPSTPADSS------AHSTPADSSAHSTPSTPVDSSA  764

Qy   341  PTTPKEPAPTTTKSAPTTKEPAPTTPKEPSPTTTKEPAPTTPKEPA----PT  396
          |   |  ||  ||  ||:| |   ||::|  |  |::|   ||:|    :||
Db   765  HGTPSTPADSSAHST---PADSSAHGTPSTPVDSSAHST---PSTPVDSSAHGTPS  817

Qy   397  TP-KKPAPTTPKEPAPTTTKKPAPTTKKPAPTTPKEPAPTTPKKL---TP  451
          || | ||  ||::|  ||  :|:||  || :|   || :||::|   ||
Db   818  TPVDSSAHSTPSTPVDSSAHGTPSTPVDSSAHSTPSTPADSSAHST-PSTPADSSAHGTP  876

Qy   452  TTP-EKLAPTTPKKPAPTTPEELAPTTPEE----PTPTTPEEPA---PTPTTPEEPA  501
          :||  :|| |    |  ||::|     ||      || ::|  |   || ::|  |
Db   877  STPVDSSAHGTPSTPADSSAHS-TPSTPVDSSAHSTPSTPADSSAHGTPSTPVDSSAHGT  935

Qy   502  PKEPAPTTPKEPAPTTPKEPAPTTLKEPAPTTPKKPAPKELAPT  561
          |  |  ||  ||  || |  ||::  |  ||::| |  ||:|  |
Db   936  PSTPADSS---AHSTPADSSAHST-------PSTPVDSSAHST  983

Qy   562  TTKEPTSTTSDKPAPTTPKGTAPTTPKE-PAPTTPKEPAPTTPKGTAPTTLKEPAPTTPK  620
          ||::|  :||   :||   ||| :||  || ::|  |  || :|||  || ||:|||  |
Db   984  -----PSTPADSSAHST----PSTPADSSAHSTPSTPVDSSAHSTPSTPVDSSAHSTPS 1033

Qy   621  KPAPKELAPTTTKGPTSTTSDKPAPTTPK----ETAPTTPKEPAPTTPKKPAPTTPETPP  676
          || :|  ||  ||||:| ::  | ||  |    | ||||||  ||:||:  ||||:|
Db  1034  TP-----ADSSAHGTPSTPADSSAHSTPSTPVDSSAHSTPSTPADSS---AHGTPSTPA 1084

Qy   677  PTTSEVSTPTTTKEPTTIHKSPDESTP-ELSAEPTPKALENSPKEPGVPTTKTPAATKPE  735
          :::  ::|||||  |:| :|:|  :::| :|  ||||:|  |  |  | ::||| |||||
Db  1085  DSSAH-STPSTPAD-SSAHGTP--STPADSSAHSTP--------------STPAGSSAN 1125

Qy   736  MT  737
          |
Db  1126  GT 1127


RESULT 10
CPN_DROME
ID   CPN_DROME      STANDARD;     PRT;    865 AA.
AC   Q02910;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   CALPHOTIN.
GN   CPN OR CAP.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=93165729; PubMed=8094559;
RA   Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT   "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=93165730; PubMed=8434015;
RA   Ballinger D.G., Xue N., Harshman K.D.;
RT   "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT   calcium and contains a leucine zipper.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC   -!- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC       REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC       OF CA+2 PER MOL OF PROTEIN.
CC   -!- SUBUNIT: HOMODIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC   -!- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC       COMPOUND EYES AND OCELLI.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC       DEVELOPMENT.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; L02111; AAA28405.1; -.
DR   EMBL; L05080; AAA28420.1; -.
DR   PIR; A47282; A47282.
DR   FlyBase; FBgn0010218; Cpn.
KW   Calcium-binding.
FT   CONFLICT     36     36       A -> AVAPAVVA (IN REF. 2).
FT   CONFLICT     43     43       I -> T (IN REF. 2).
FT   CONFLICT     64     64       I -> V (IN REF. 2).
FT   CONFLICT     76     76       T -> A (IN REF. 2).
FT   CONFLICT    100    100       P -> PP (IN REF. 2).
FT   CONFLICT    126    127       VQ -> AP (IN REF. 2).
FT   CONFLICT    154    154       I -> V (IN REF. 2).
FT   CONFLICT    160    160       S -> T (IN REF. 2).
FT   CONFLICT    534    534       A -> E (IN REF. 2).
FT   CONFLICT    699    699       I -> T (IN REF. 2).
FT   CONFLICT    703    703       V -> L (IN REF. 2).
FT   CONFLICT    721    721       D -> E (IN REF. 2).
```

Mon Apr 29 08:35:15 2002

DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCFP1).
GN   FP1.
OS   Mytilus coruscus (Sea mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=42192;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foot;
RX   MEDLINE=96394686; PubMed=8798340;
RA   Inoue K., Takeuchi Y., Takeyama S., Yamaha E., Yamazaki F., Odo S.,
RA   Harayama S.;
RT   "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and
RT   its evolutionary implications.";
RL   J. Mol. Evol. 43:348-356(1996).
CC   -!- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC   PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC   ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC   FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC   -!- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC   -!- PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND
CC   ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; D63777; BAA09850.1; -.
DR   InterPro; IPR002964; Adhesive_plaq.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01216; ADHESIVEI.
DR   PRINTS; PR01217; PRICHEXTENSN.
KW   Signal; Repeat; Hydroxylation.
FT   SIGNAL   1   20   POTENTIAL.
FT   CHAIN   21   872   ADHESIVE PLAQUE MATRIX PROTEIN.
FT   DOMAIN   21   41   NONREPETIVE LINKER.
FT   DOMAIN   124   872   TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-
FT   P-[PST]-[ST].
FT   DOMAIN   184   192   NONAPEPTIDE 1.
FT   DOMAIN   213   221   NONAPEPTIDE 2.
SQ   SEQUENCE   872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;

Query Match   7.9%; Score 519; DB 1; Length 872;
Best Local Similarity 28.9%; Pred. No. 5.7e-15;
Matches 282;   Conservative 82;   Mismatches 420;   Indels 192;   Gaps 55;

Qy   11 LLLLSVFVIQQVSSVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTST-TQH 68
           | ||:| || | : :::| | || || ::|| ::|
Db   8 LCLLCIFSCDIFALSNGNIHN-----------------VYGSAYSGASAGAYK-TLPGSHPYGSKH 55

Qy   69 ------NKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQT 121
                 ::|| | : | | || || | || : | | || | |
Db   56 VPVYKPMNKIPT-PYI--SKKSYPAPYKPKGYPTKRYQPTYGSKTNYPPIYKPIAKKLS 112

Qy   122 STDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPT-PTTP 180
                    | | | || || | : | || ||| | | | | |||
Db   113 SYKAIKTTYPAYKAKTSYPPSYK-----HKITYPPIYK------PKITYP--PIYKQKP 158

Qy   181 KEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPT-TPKEPAPT--TTKPAPTTPKEP 237
           | ||| || || || :||| | |: | | ||||| ||||| || |
Db   159 SYPPSYKPKTTYPPTYK-----PKITYPPTYKRKPSTTPYKPKATYPPTYKPKITYP--- 210

Qy   238 APTTTKEPAPTTTKSAPTTPKEPAPTT--PKKPAPTTPKEPAPTT----PKEPTPTT--P 289
           || ||| || || || | | || || | || | ||| | | |

```
Db    211  -PTYKRKPSYT-----PYKPKTYPPTYKPKISYPSIYKPKASYVSSYKKKTYPPTYKP    264
                :||  :||:|:  ||
Qy    290  KEPAPTTKEPAPTPKEPAPT-APKKPAPTTPKEPAPTTPKEPAPTTKEPSPTTPKEPA    348
                :||::||:|
Db    265  KISYPPTYKPKPSYP----PTYKPKVTYPPTYK-PKPSYPPTYKPKITYPPT-YKPKPSY  318

Qy    349  PTTTKSAPT-----TTKEPAPTTKSAPTTPKEPSPTTKEPA-PTTPKEPAPTTPKKP-    401
                ||  |:  :|  || ||  :|
Db    319  PTPYKQKPSYPIYKSKSSYPTSYKSKKKTYPPTYKPKITYPPTYKPKPSYPPSYKPKKTY  378

Qy    402  APT-TPK-------EPAPTTPKEPAPTTTKPAPTAPKEPAPTTKETAP-------TTPKK  448
                :|| ||         || |||
Db    379  SPTYKPKITYPPTYKPKPSYPSYKPKTTYPPT-YKPKISYPPTYKPKASYVSSYKSKKT   437

Qy    449  LTPTTPEKLAPTTPEKAPTTPEELAPTTPEPT-PTTPEEPAPTTPKAAAPNTPK-EPA    506
                ||   |:||||  :|:|
Db    438  YPPTYKPKISYPPTYKPKSYPPTYKPKITYPPTYKPKPSYPSYPPTYKPKITYPPTYKRKPS 497

Qy    507  PTTPKEPAPTTP------KEPAPT--KEPAPTT-PKGTAPTTLKEPAPTTPKKPAPKE    557
                ||||| || ||
Db    498  YPPTYKQKPSYPPIYKSKSSYPTSYKSKKTYPPTYKPKITYPPTYK------PKPSYPPS   551

Qy    558  LAPTTTKEPTSTTSDKPAPT-TPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAP  616
                || |
Db    552  YKPKTYPPTYKPKIRYPPTYKPKASYPPTYK------------PKITYPPTYK-PKP    596

Qy    617  TTP----KKPAPKELAPTTTKGPTSTTSDKPAPTT--PKETAPTTPKEPAPTTPK--KPA  668
                ||  :|
Db    597  SYPTPYKQKPSYPPIYKSKSSYPTAVKSKKTYPPTYKPKITYPPTYK-PKPSYPPSYRPK   655

Qy    669  PTTPET-------PPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEP  721
                || |
Db    656  ITYPPTYKPKKSYPQAYKSKGSYPPSYQPKKTY----PPSYKPKKTYPPTYK-----PKIS  707

Qy    722  GVPTTKTPAATKPEMTTAKDKT----TERDLRTPETTTAAPKMTKETATTEKTTESK    777
                || :|
Db    708  YPPTYK----TKPSYPASYKRKTSYPPTYKPKISYPSTYKAKP-----SYPPTYKPKPSY   758

Qy    778  ITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRAT  837
                 ||
Db    759  ASSYKPKIRYPPTYKPKPSYASSYK-----PKI---------RYPPTYKPKP---      796

Qy    838  NSKATTPKPQ---KPTKAPKKP-TSTKKPK-TMPRVRKPKTT--PTPRKMTSTMPELNPT  890
                 |:
Db    797  -SYASSYKPKIRYPPTYKPKPSYASSYKPKITYPPTYKPKISYPPTYKPKITYPPTYKPK  855

Qy    891  SRIAEAMLQTTTRPNQ  906
                 ::|
Db    856  ISYPPAYKPKISYPSQ  871
```

```
CC   -!- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC       INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC       MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SKIN.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   -!- SIMILARITY: CONTAINS 6 P-TYPE (TREFOIL) DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L02115; AAA74725.1; -.
DR   PIR; A45155; A45155.
DR   HSSP; P04002; 1WFA.
DR   InterPro; IPR000519; P_trefoil.
DR   Pfam; PF00088; trefoil; 6.
DR   SMART; SM00018; P; 6.
DR   PROSITE; PS00025; P_TREFOIL; 6.
KW   Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
FT   NON_TER         1      1
FT   DOMAIN         81    144       8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT                                  ALA/THR-RICH.
FT   REPEAT         81     88       1-1.
FT   REPEAT         89     96       1-2.
FT   REPEAT         97    104       1-3.
FT   REPEAT        105    112       1-4.
FT   REPEAT        113    120       1-5.
FT   REPEAT        121    128       1-6.
FT   REPEAT        129    136       1-7.
FT   REPEAT        137    144       1-8.
FT   DOMAIN        161    202       P-TYPE 1.
FT   DOMAIN        218    301       8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT   REPEAT        218    224       2-1.
FT   REPEAT        225    239       2-2.
FT   REPEAT        240    249       2-3.
FT   REPEAT        250    259       2-4.
FT   REPEAT        260    275       2-5.
FT   REPEAT        276    287       2-6.
FT   REPEAT        288    294       2-7.
FT   REPEAT        295    301       2-8.
FT   DOMAIN        306    347       P-TYPE 2.
FT   DOMAIN        353    394       P-TYPE 3.
FT   DOMAIN        402    522       12 X APPROXIMATE TANDEM REPEATS,
FT                                  THR-RICH.
FT   REPEAT        402    411       3-1.
FT   REPEAT        412    419       3-2.
FT   REPEAT        420    431       3-3.
FT   REPEAT        432    443       3-4.
FT   REPEAT        444    453       3-5.
FT   REPEAT        454    460       3-6.
FT   REPEAT        461    472       3-7.
FT   REPEAT        473    479       3-8.
FT   REPEAT        480    491       3-9.
FT   REPEAT        492    498       3-10.
FT   REPEAT        499    515       3-11.
FT   REPEAT        516    522       3-12.
FT   DOMAIN        525    566       P-TYPE 4.
FT   DOMAIN        572    613       P-TYPE 5.
FT   DOMAIN        620    661       P-TYPE 6.
FT   DISULFID      162    188       BY SIMILARITY.
FT   DISULFID      172    187       BY SIMILARITY.
FT   DISULFID      182    199       BY SIMILARITY.
FT   DISULFID      307    333       BY SIMILARITY.
FT   DISULFID      317    332       BY SIMILARITY.
FT   DISULFID      327    344       BY SIMILARITY.
FT   DISULFID      354    380       BY SIMILARITY.
FT   DISULFID      364    379       BY SIMILARITY.
```

```
RESULT  12
MUC1_XENLA
ID   MUC1_XENLA     STANDARD;      PRT;     662 AA.
AC   Q05049;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   INTEGUMENTARY MUCIN C.1 (FIM-C.1) (FRAGMENT).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=93077556; PubMed=1447205;
RA   Hauser F., Hoffmann W.;
RT   "P-domains as shuffled cysteine-rich modules in integumentary mucin
RT   C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic
RT   polymorphism.";
RL   J. Biol. Chem. 267:24620-24624(1992).
```

```
FT   DISULFID    374    391             BY SIMILARITY.
FT   DISULFID    526    552             BY SIMILARITY.
FT   DISULFID    536    551             BY SIMILARITY.
FT   DISULFID    546    563             BY SIMILARITY.
FT   DISULFID    573    599             BY SIMILARITY.
FT   DISULFID    583    598             BY SIMILARITY.
FT   DISULFID    593    610             BY SIMILARITY.
FT   DISULFID    621    647             BY SIMILARITY.
FT   DISULFID    631    646             BY SIMILARITY.
FT   DISULFID    641    658             BY SIMILARITY.
FT   VARIANT     276    276             K -> E.
FT   VARIANT     354    354             C -> R.
FT   VARIANT     415    415             T -> A.
SQ   SEQUENCE   662 AA;  67774 MW;  F085277F1ED2FD40 CRC64;
```

```
Db        567 LPQVADCKVAPSSR 580
                  :|:| | |:|

RESULT  13
ID   RPB1_HUMAN     STANDARD;      PRT;   1970 AA.
AC   P24928;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).
GN   POLR2A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92178992; PubMed=1542581;
RA   Wintzerith M., Acker J., Vicaire S., Vigneron M., Kedinger C.;
RT   "Complete sequence of the human RNA polymerase II largest subunit.";
RL   Nucleic Acids Res. 20:910-910(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95347616; PubMed=7622068;
RA   Mita K., Tsuji H., Morimyo M., Takahashi E., Nenoi M.,
RA   Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
RT   "The human gene encoding the largest subunit of RNA polymerase II.";
RL   Gene 159:285-286(1995).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC       RNA(N).
CC   -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC       THE PHOSPHORYLATION ACTIVATES POL2.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC       FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC       PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC       III FOR 5S AND TRNA GENES.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X63564; CAA45125.1; -.
DR   EMBL; X74874; CAA52862.1; -.
DR   EMBL; X74873; CAA52862.1; JOINED.
DR   EMBL; X74872; CAA52862.1; JOINED.
DR   EMBL; X74871; CAA52862.1; JOINED.
DR   EMBL; X74870; CAA52862.1; JOINED.
DR   PIR; S21054; S21054.
DR   MIM; 180660; -.
DR   InterPro; IPR000684; RNA_polII_repeat.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR002879; RNA_pol_A2.
DR   Pfam; PF00623; RNA_pol_A; 1.
DR   Pfam; PF01854; RNA_pol_A2; 1.
DR   PROSITE; PS00115; RNA_POL_II_REPEAT; 43.
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW   DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT   ZN_FING      71     87       C2H2-TYPE (POTENTIAL).
FT   DOMAIN     1590   1958       CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT   CONFLICT   1067   1067       W -> L (IN REF. 2).
FT   CONFLICT   1449   1449       D -> Y (IN REF. 2).
```

```
Query Match           7.8%;   Score 509.5;  DB 1;  Length 662;
Best Local Similarity   28.1%;  Pred. No. 1.1e-14;
Matches  223;  Conservative  48;  Mismatches  276;  Indels  247;  Gaps  30;

Qy   116 TTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEP 175
             ||  |:| :|::. : |||:|                 |||
Db     3 TTAAVAATGKDTTAAAEGSAAAEKTAA-----AGEVSAPPT--AAVAATGEDATT----  51

Qy   176 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTKSAPTTPKEPAPTTKEPAPTTPK 235
             |||    :|  |||| ||| |||   || |       |||
Db    52 ------AAATAAAETTAAAGEAPTTTTAPATTAAGKAPTTAAATAPTTAAAGAPTTAT 103

Qy   236 EPAPTTTKEPAPTTKS-APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPKEPA- 293
             || ||  |  |||| |      :   :    |||  : |   : .|
Db   104 GKAPATAAAPVPTTAASKAPTTAAAAPTTAASAAKSKERSTSSSSEEEHCH 163

Qy   294 --PTTKEPAPT---TPKEPAPTAPKKPAPTTPK------EPAPTTPKEPAPTTTKE 338
             |:|             |||  ||
Db   164 VKPSKRMCGSKGITKKQ-----CKKKNCCFDPKGHGGIHCFHRKPKGHSHEEHTTTTK- 218

Qy   339 PSPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTP 398
             |||| ||  |||| |||           |||
Db   219 ----APTTIQIATTTT---TPTTT------------TTTTKATPTT-------------- 245

Qy   399 KKPAPTTPKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLIPTTPEKLA 458
             ||||        || |||| ||  |||    ||||
Db   246 -------------------TTTKATPTTT---------TTTKATTTTT----TPTT----- 270

Qy   459 PTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTP 518
             |||   |  ||       |||  || |||
Db   271 --------TTTTKATTPTTTTTTTP---------------TTTTTKATPTT---------- 289

Qy   519 KEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKEPTSTTSDKPAPTT 578
             |||||  |    || |||       :|      |  ||||
Db   290 -------TTTTKATTTTTTTSGECKMEPSK---------REDCGYSGITESQCR 328

Qy   579 PKG-----TAPTT-----PKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELA 628
             ||:          ||||
Db   329 TKGCCFDSSIPQTKWCFYTLSQVADCKVEPSQRVDCGFRGIT----ADQCRQKNCCFDSS 384

Qy   629 PTTTKGPISTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPEIPPPTTSEVSTPTTT 688
             ||||  | |||||      ||      ||      ||||
Db   385 ISGTKWCFYSTSQVAA--TKTTTTTPTTTTPTTTTTKATTTTP------TTTTTPTTT 436

Qy   689 KEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERD 748
             |||| ||||                            ||||
Db   437 TTTTT-----TTKA---------TTTTTTPTTTTTTT---KAT----KAT--- 463

Qy   749 LRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTTSTTTQDTTPFKITTLKTTTLAP 808
             |||| ||       :|||: |  |||       ||||
Db   464 -TTTPTTTTTTP----TTTTTKAT------TTTPTTTTTTTTPTTTTTKATTTTPTTTTT 510

Qy   809 KVTTTKKTTTTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPK----T 864
             ||||| ||||                    :||  :|:|
Db   511 TTTTTKATTTTTGECKME---PSKRADCGYPGITESQCRSKGCCFDSSIPQTKWCFYS 566

Qy   865 MPRVRKPKTTPTPR 878
```

```
SQ   SEQUENCE   1970 AA;  217205 MW;  6876FC25692A657E CRC64;

Query Match
Best Local Similarity    7.7%;   Score 503.5;  DB 1;  Length 1970;
Matches 181;  Conservative 101;  Mismatches 167;  Indels 95;  Gaps    47;

Qy    175 PTPTTPKEPASTTPKE-PTPTTIKSAPTTPKEPAPTTTKSAPTTPKE-----PAPTTTKE 228
              ::|| |  ||  |            ||   :|:    ||::|
Db   1507 PSPMGGISPAMTPWNQGATPAYGAWSPSVGSGMTPGAAGFSPSAASDASGFSPGYSPAWS 1566

Qy    229 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEP--APTTPKEPTP 286
              |  | | :|:|  | ||         || | |  |
Db   1567 PTPGSPGSPGPSSPYIPSPGGAMSPSYSPTSPA-YEPRSPGGYTPQSPSYSPTSPSY-SP 1624

Qy    287 TTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTTKEPSPTTPKE 346
              |:| :|| :|:|  |       ||    |||:
Db   1625 TSPSY-SPTSPNYSPTSPSY-SPTSPSY-SPTSPSY-------SPTSPSY 1671

Qy    347 PAPTTTKSAPTTTKEPAPTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTTP 406
              :|:   |||  :|:  :|:|   |: | |  ||::    ||:|   :||:|
Db   1672 -SPTSPSYSPTSPSY-SPTSPSY-SPTSPSY -SPTSPSY-SPTSPSY-SPTSP 1725

Qy    407 KEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPA 466
              :|||   :||    :||::    |::|    ||:   |  :||
Db   1726 SY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPNY--PTSP-SYSPTSP-SYSPTSPNY-T 1777

Qy    467 PTTPEELAPTTPEEPTPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEP--APT 524
              ||:|   ||| |       :::|:     |: | |  ||:|:|
Db   1778 PTSP-NYSPTSPSY-SPTSP-SYSPSPR----YTPQSPT-YTPSSPSYSPS 1828

Qy    525 TPKETAPTTPKGTAPTTLKEPAPTTPK-KPAPKELAPTTTKEPTSTTSDKPAPTTPKGTA 583
              :| :|:|:|      |  :|||     :|:|:|   ||::||::
Db   1829 SP-SYSPTSPKYT--------PTSPSYSPSSPEYTPTSPK--YSPTSPKYSPTSPK-YS 1875

Qy    584 PTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPT-STTSDK 642
              ||:|  ||  |  :|:|    |||      :||     |||  |  ||:  :||
Db   1876 PTSPTY-SPTTPKY-SPTSP--TYSPTSPVYTPTSPK-------YSPTS--PTYSPTSPK 1922

Qy    643 PAPTTPKETAPTTPKEP--APTTPKKPAPTTPTEPTPPPTTSEVSTPTTTKEPTTIHKSPDE 700
              :|  |::|   |  :||     :|  |||||    ||||           | | ||::||
Db   1923 YSPTSP-TYSPTSPKGSYSPTSPGY-SPTSP------TYSLTSPAI--------SPDD 1965

Qy    701 STPE 704
              |  |
Db   1966 SDEE 1969
```

RESULT 14
RPB1_MOUSE
ID   RPB1_MOUSE   STANDARD;   PRT; 1970 AA.
AC   P08775;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).
GN   POLR2A OR RPO2-1 OR RPII215.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87280135; PubMed=3038894;
RA   Ahearn J.M. Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.;
RT   "Cloning and sequence analysis of the mouse genomic locus encoding
RT   the largest subunit of RNA polymerase II.";
RL   J. Biol. Chem. 262:10695-10705(1987).
RN   [2]
RP   SEQUENCE OF 1587-1970 FROM N.A.
RX   MEDLINE=86068017; PubMed=2999785;
RA   Corden J.L., Cadena D.L., Ahearn J.M. Jr., Dahmus M.E.;
RT   "A unique structure at the carboxyl terminus of the largest subunit
RT   of eukaryotic RNA polymerase II.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).
RN   [3]
RP   REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.
RX   MEDLINE=92178992; PubMed=1542581;
RA   Wintzerith M., Acker J., Vicaire S., Vigneron M., Kedinger C.;
RT   "Complete sequence of the human RNA polymerase II largest subunit.";
RL   Nucleic Acids Res. 20:910-910(1992)
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC   OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC   SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC   RNA(N).
CC   -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC   THE PHOSPHORYLATION ACTIVATES POL2.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC   FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC   PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC   III FOR 5S AND TRNA GENES.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M12130; AAA40071.1; -.
DR   EMBL; M14101; AAA40071.1; JOINED.
DR   PIR; A28490; A28490.
DR   MGD; MGI:98086; Rpo2-1
DR   InterPro; IPR000684; RNA_polII_repeat.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR002879; RNA_pol_A2.
DR   Pfam; PF00623; RNA_pol_A; 1.
DR   Pfam; PF01854; RNA_pol_A2; 1.
DR   PROSITE; PS00115; RNA_POL_II_REPEAT; 42.
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW   DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT   ZN_FING   71   87   C2H2-TYPE (POTENTIAL)
FT   DOMAIN   1590   1958   CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT   CONFLICT   1498   1498   P -> R (IN REF. 1 AND 2).
FT   CONFLICT   1499   1536   MISSING (IN REF. 1 AND 2).
SQ   SEQUENCE   1970 AA; 217175 MW; 7D76F38FD92A657E CRC64;

Query Match
Best Local Similarity   7.8%;   Score 498.5; DB 1; Length 1970;
Matches 180; Conservative 101; Mismatches 168; Indels 95; Gaps   47;

Qy   175 PTPTTPKEPASTTPKE-PTPTTIKSAPTTPKEPAPTTTKSAPTTPKE-----PAPTTTKE 228
              ::|| | || | || :|: ||::|
Db   1507 PSPMGGISPAMTPWNQGATPAYGAWSPSVGSGMTPGAAGFSPSAASDASGFSPGYSPAWS 1566

Qy   229 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTPKEPAPTTPKKPAPTTPKEP--APTTPKEPTP 286
              | | | :|:| | || || | | |
Db   1567 PTPGSPGSPGPSSPYIPSPGGAMSPSYSPTSPA-YEPRSPGGYTPQSPSYSPTSPSY-SP 1624

Qy   287 TTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTTKEPSPTTPKE 346
              |:| :|| :|:| | || |||:
Db   1625 TSPSY-SPTSPNYSPTSPSY-SPTSPSY-SPTSPSY-------SPTSPSY 1671

Qy   347 PAPTTTKSAPTTTKEPAPTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTTP 406
              :|: ||| :|: :|:| |: | | ||:: ||:| :||:|
Db   1672 -SPTSPSYSPTSPSY-SPTSPSY-SPTSPSY -SPTSPSY-SPTSPSY-SPTSP 1725

Qy   407 KEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPA 466
              :||| :|| :||:: |::| ||: | :||
Db   1726 SY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPNY--PTSP-SYSPTSP-SYSPTSPNY-T 1777
```

```
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC       ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M37227; AAA41693.1; ALT_FRAME.
DR   EMBL; X13804; CAA32038.1; ALT_FRAME.
DR   EMBL; M21964; AAA41695.1; -.
DR   EMBL; J04517; AAA41692.1; -.
DR   PIR; A30796; A30796.
DR   PIR; A25649; A25649.
DR   PIR; B25649; B25649.
DR   PIR; S02003; S02003.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
FT   NON_TER        1      1
FT   DOMAIN       276    641       51 X 3 AA TANDEM REPEATS OF K-S-P.
FT   CONFLICT     164    164       L -> I (IN REF. 2).
FT   CONFLICT     185    185       L -> S (IN REF. 2).
FT   CONFLICT     193    193       L -> T (IN REF. 2).
FT   CONFLICT     199    199       M -> T (IN REF. 2).
FT   CONFLICT     346    346       K -> N (IN REF. 1 AND 4).
FT   CONFLICT     373    373       A -> V (IN REF. 1 AND 4).
FT   CONFLICT     482    482       G -> E (IN REF. 2 AND 4).
FT   CONFLICT     485    485       P -> S (IN REF. 2).
FT   CONFLICT     570    571       RK -> KE (IN REF. 2 AND 4).
FT   CONFLICT     591    591       P -> T (IN REF. 2 AND 4).
FT   CONFLICT     727    727       A -> V (IN REF. 4).
FT   CONFLICT     757    759       AAP -> GST (IN REF. 4).
FT   CONFLICT     769    769       T -> L (IN REF. 2).
FT   CONFLICT     775    775       R -> P (IN REF. 2 AND 4).
SQ   SEQUENCE   831 AA;  89486 MW;  1B0973C3F13EF768 CRC64;

Query Match                   7.6%;  Score 497;  DB 1;  Length 831;
Best Local Similarity        27.6%;  Pred. No. 4.3e-14;
Matches  195;  Conservative  59;  Mismatches  307;  Indels  146;  Gaps  35;

Qy      61  PDTSTTQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTNKQ  120

Db     197  PSMST--HIKVKSEEKIKVVE---------------KSEKETVIVEEQTEEIQVTEEVTEEED  242

Qy     121  TSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKEP---T  176
                  :|:   |  |  ||        |: ||  ||         ||| |: |
Db     243  KEAQGEEEEAEEGGEEAATTS-----PPAEEAASP---EKETKSPVKEEAKSPAEAKS  293

Qy     177  PTTPKEPA-STTPKEPTPTTIKSAPTTPKEPA------PTTTKSAPTTPKEPA----PTTTK  227
                  | ||| :|| :||:           |:           :|  ||  |  ||
Db     294  PAEAKSPAEAKSPAEVKSPAVAKSPAEVKSPAEAKS-PAEAKSPAEVKSPATVK  352

Qy     228  EP----APTTPKEPA-----PTTTKEPA----PTTTKEPA----PTTPKKPA-  269
                      | |       | ||      :|  ||       : ||
Db     353  SPGEAKSPAEAKSPAEVKSPAEAKSPAEAKSPASVKS--PGEAKSPAEAKSPAEVKSPATV  411

Qy     270  --PTTPKEPA----PTTPKEP---TPTTPKEPAPTTKEP----APTTPKEPA----PTA  311
```

aal.rsp

```
Db   412 KSPVEAKSPAEVKSPVTVKSPAEAKSPVEVKSPA-SVKSPSEAKSPAGAKSPAEAKSPVV 470
             |  |||    ||||  |||   |  |   ||    :||  :| || :  |||   |
Qy   312 PKKPAPTTPKEPAPTTPKEPSPTTPKEPA----PTTTKS----APTTTKEP 362
             ||  ||   :   |  : ||   :   ||        |        :|  |
Db   471 AKSPAEA--KSPAGAKPPAEAKSPAEAKSPAEAKSPAEAKSPVEVKSPEKAKSP 528
             |||    |||     |  ||   |  ||   |   ||    |||   |
Qy   363 APTTTKS-----APTTPKEPSPTTTKEPA----PTTPKEPAPTTPKKPAPTTPKEPAPTT 413
               :|      |  | |   :  ||                    :|   |
Db   529 VKEGAKSLAEAKSPEKAKSPVKEEIKPPAEVKSPEKAKSPMRKEAKSPEKAKTLDVKSPE 588
             |   |||   ||   |  ||                 |
Qy   414 PKEPAPTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEEL 473
             |   | |   ||| |     |   | |  : |:||  | :|||   |:|  ||:  ||:
Db   589 AKPPAKEEAKRPADIRSPEQVKSPAKEEAKSPEKEETRT--EKVAPKKEEVKSPV--EEV 644
             |      |  |       |   |  |   |  ||      ||    :   ||   |||
Qy   474 APTTPEEPTPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTT 533
             :||    | |||    |  | ||  |     |:|
Db   645 --KAKEPPKKVEEKTPATPKTEVKESKKDEAPKEAQKP-KAEEKEPLTEKPKDSPGEA 700
             | ||      ||       |    ||   |:   ||     ||| |:
Qy   534 PKGTAPTTLKEPAPTTPKKPAPKELAPTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPT 593
             | ||  |     |  |::    |   |  |    :   |  | |||  ||   |
Db   701 KKEEA----KEKKAAAPEEETPAKLGVKEEEAKPKEKAEDAKA-----------KEPSKP 744
                :|   ::  |    |      |:   || |     |:
Qy   594 TPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAP 653
             :|||  ||     ||  |:  |:  :|  |  ||
Db   745 SEKE----KPK------KEEVPAAPEKKDYKE--------EKTTESKKREEKPKMEAK 784
             |  |   |:||           |   | ||
Qy   654 TTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDE 700
             ::         |: |: || :: :   |:|:  |:|:|:
Db   785 AKEED------KGLPQEPSKPKTEKAEKSSSTDQKDSQPSEKAPED 824
```

Search completed: April 26, 2002, 16:32:03
Job time: 553 sec
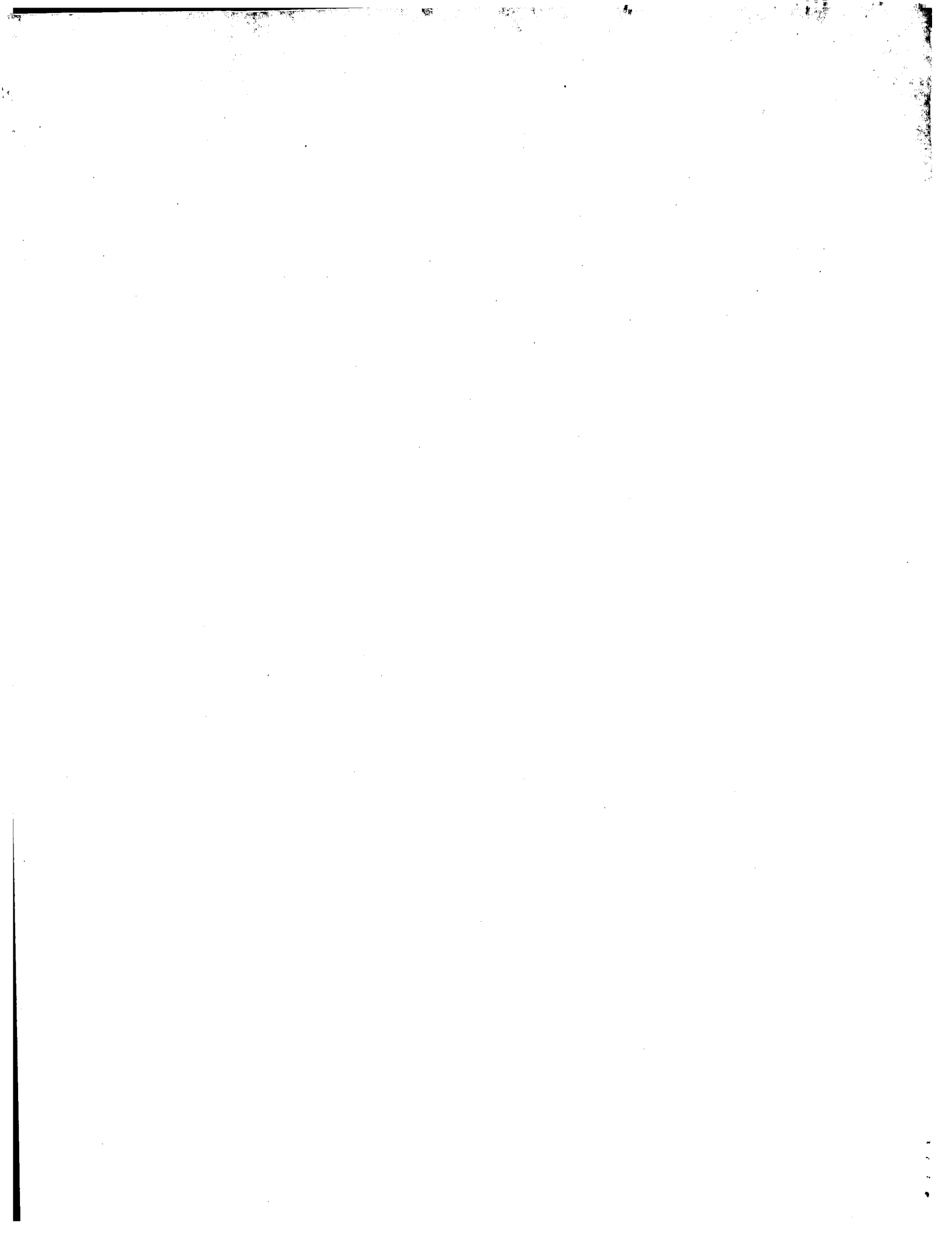
```
                    GenCore version 4.5
              Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:       April 26, 2002, 16:17:40 ; Search time 61.21 Seconds
                                          (without alignments)
                                          451.831 Million cell updates/sec

Title:          AA1
Perfect score:  6568
Sequence:       1 MAWKTLPIYLLLLSVFVIQ.........ARAITTRSGQTLSKVWNCP 1229

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters:       212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Issued_Patents_AA:*
          1:    /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
          2:    /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
          3:    /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
          4:    /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
          5:    /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
          6:    /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	522	7.9	805	4	US-09-103-429A-4	Sequence 4, Appli
2	508	7.7	1837	5	US-08-928-361B-5	Sequence 5, Appli
3	506.5	7.7	744	6	5202236-25	Patent No. 5202236
4	498.5	7.6	786	4	US-09-103-429A-3	Sequence 3, Appli
5	489	7.4	1721	3	US-08-700-651-5	Sequence 5, Appli
6	489	7.4	1721	3	US-08-928-361B-6	Sequence 6, Appli
7	488.5	7.4	826	1	US-07-638-431-2	Sequence 2, Appli
8	488.5	7.4	826	6	PCT-US92-00018-2	Sequence 2, Appli
9	476.5	7.3	652	2	5202236-13	Patent No. 5202236
10	452	6.9	960	4	US-09-219-849-5	Sequence 5, Appli
11	424.5	6.5	1867	4	US-08-479-537A-5	Sequence 5, Appli
12	424.5	6.5	1867	1	US-08-083-116-5	Sequence 53, Appl
13	424.5	6.5	2035	2	US-08-479-537A-2	Sequence 68, Appl
14	424.5	6.5	2035	2	US-09-083-116-2	Sequence 2, Appli
15	419.5	6.4	2476	2	US-08-276-967-2	Sequence 2, Appli
16	419	6.4	1185	4	US-09-041-886-23	Sequence 23, Appl
17	417	6.3	829	4	US-08-642-255-132	Sequence 132, App
18	417	6.3	829	1	US-08-397-633A-53	Sequence 53, Appl
19	417	6.3	837	1	US-08-175-155-68	Sequence 68, Appl
20	417	6.3	837	1	US-08-477-509B-103	Sequence 103, App
21	417	6.3	837	1	US-08-642-255-101	Sequence 101, App
22	417	6.3	837	2	US-08-707-237A-75	Sequence 75, App
23	417	6.3	837	1	US-08-477-509B-103	Sequence 103, App
24	417	6.3	897	1	US-08-397-633A-50	Sequence 50, Appl
25	413.5	6.3	907	3	US-08-783-774-2	Sequence 2, Appli
26	413.5	6.3	907	1	PCT-US95-04611A-19	Sequence 19, Appl
27	404.5	6.2	408	1	US-07-609-716-65	Sequence 65, Appl
28	404.5	6.2	408	4	US-08-475-411A-65	Sequence 65, Appl
29	404.5	6.2	408	4	US-08-478-029A-65	Sequence 65, Appl
30	398.5	6.1	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	6.1	682	1	US-08-397-633A-36	Sequence 36, Appl
32	381	5.8	1848	4	US-08-296-791-6	Sequence 6, Appli
33	381	5.8	1848	5	PCT-US95-1061A-6	Sequence 6, Appli
34	376	5.7	960	4	US-09-219-849-6	Sequence 6, Appli
35	374.5	5.7	1537	2	US-08-325-267A-2	Sequence 2, Appli
36	373.5	5.7	1231	3	US-08-904-263A-4	Sequence 4, Appli
37	371	5.6	761	2	US-08-707-237A-84	Sequence 84, Appl
38	371	5.6	762	1	US-08-642-255-114	Sequence 114, App
39	371	5.6	762	1	US-08-397-633A-26	Sequence 26, Appl
40	370.5	5.6	1064	1	US-08-642-255-62	Sequence 62, Appl
41	368	5.6	762	1	US-08-642-255-120	Sequence 120, App
42	368	5.6	762	1	US-08-397-633A-31	Sequence 31, Appl
43	367.5	5.6	1187	3	US-08-320-559-28	Sequence 28, Appl
44	367.5	5.6	1187	3	US-08-545-860D-28	Sequence 28, Appl
45	367.5	5.6	1187	5	PCT-US94-04496-28	Sequence 28, Appl

ALIGNMENTS

```
RESULT 1
US-09-103-429A-4
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
;  APPLICANT: Granados, Robert R
;  APPLICANT: Wang, Ping
;  TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
;  TITLE OF INVENTION: cDNA and Related Products and Methods
;  NUMBER OF SEQUENCES: 4
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
;   STREET: 118 No. 6187558th Tioga
;   CITY: Ithaca
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 14850
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/103,429A
;   FILING DATE: 24-JUN-1998
;   CLASSIFICATION: 800
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Michaels, Christopher A
;   REGISTRATION NUMBER: 34,390
;   REFERENCE/DOCKET NUMBER: BTI-39
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (607) 256-2000
;   TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 805 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;  MOLECULE TYPE: protein
;  HYPOTHETICAL: NO
;  ANTI-SENSE: NO
;  ORIGINAL SOURCE:
;   ORGANISM: Trichoplusia ni
;   TISSUE TYPE: peritrophic membrane
US-09-103-429A-4

Query Match             7.9%;  Score 522; DB 4;  Length 805;
Best Local Similarity   27.8%;  Pred. No. 1.9e-27;
```

```
Matches 226;  Conservative   33;  Mismatches  211;  Indels  342;  Gaps   40;

Qy    149  TSKVLAKPTPKAETTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAP-----TTP  203
                :: :| | |||| || |:|    |  ||    :|
Db      9  TALGLVAARPEVSDAEKNPALHEPHPDXP--PAEQXXILPXEYDCTKFYYCEYGLKFIAP   66

Qy    204  KEPAPTT-----------TKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTK  251
             :: ||| :          :  | | ||| :   | ||||: |||||
Db     67  RDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETT----QAPATTQ--APTTTQ--APTTTT  119

Qy    252  SAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTA  311
             |||     |||          ||||: :| || |
Db    120  QAPTTTTQATTTT---------QAPTTTQAPTTT----QATTTQAPTTT---  156

Qy    312  PKKPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAP  371
             |||| |||||       |||||: ||: ||: |||||: |||||||
Db    157  ---QAPTTTTQ-APTTTQ-APTTTQ---------APTTTQ-APTTTQ--APTTITQAA   198

Qy    372  TTPKEPSPTTTKEPAPTTP--KEPAPTTPKKPAPTTPKEPAPTTPK-------------  415
             |||     :| |  | ||| |||  || ||| ||
Db    199  TTPAATTPAAT-TPAATTPAATTPAATTPGVPAPTS----APVWPPICELLPNGCPADFD  253

Qy    416  ------------------------------------------EPAPTTTK  423
                                                       ||| :
Db    254  IHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTE  313

Qy    424  K----------------------------------------PA------------  426
             ||                                     ||
Db    314  GNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQ  373

Qy    427  -------------------------------PTAP--KEPAPT-----------TPKET  441
                                            | || || ||           :| :
Db    374  QCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWPLPNGCPADFSIDHLLPHES  433

Qy    442  -------------------------APTTPKKLTPTT---------PEKLAPTTPEKPAP  467
                                      :|  |||||||         |:| | ||||||
Db    434  DCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPSVTAGCQVFECDSDNQCTSAAPTAAP  493

Qy    468  TTPEELAPTTPEEPTTPEEPAPTTPEEPAPNTPKEPAPTTPKEPAPTTPK  527
             ||| | |  |||   |  |    | | ||||
Db    468  TTPEELAPTTPEEPTTPEEPAPTTPEEPAPNTPKEPAPTTPKEPAPTTPK  527
```

```
Query Match            7.7%;  Score 508;  DB 3;  Length 1837;
Best Local Similarity  21.3%;  Pred. No. 4.5e-26;
Matches  295;  Conservative 100;  Mismatches  494;  Indels  496;  Gaps   48;

Qy     57  KVTTPDTSTTQHNKVSTSPKITTAKPINPRSLPPNSDTSKETSLTTVNKETTVETKETTT  116
            || ||| :::                ||| |  |||   | | ||||| :
Db    307  KHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  366

Qy    117  TNKQTSTDGKEKTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPT  176
            |:: ||: |  :  ::  |    |||
Db    367  TTTTTTTTTTKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTT  426

Qy    177  PTTPKEPASTTPKEPTPTTIKSAPTTPKEP--APTTTKSAPTTPKEPAPTTTKEPAPTTP  234
            ||| | |         || :  ||||:  ||
Db    427  TTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  486

Qy    235  KEPAPTTTKEPAPTTTKKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKE---------  283
            |||||        |||| : |||     ||
Db    487  TTTATTTTKKPTTTTTTTTTTKKPTTTTTTATTTTTSETESVIKPDEWCWLEKNGECE  546

Qy    284  --------------------------------PTPTT----------------PK  290
                                            ||||
Db    547  AKGATYVGIGKDGRIENGMAFTMIPNDDTHVRFRFKVKDVGNTISVRCGKGAGKLEFPD  606

Qy    291  EPAPTTKEP---------------------APTTPKEPA-PTAPKKP------------  315
            |||||||                      ||||||
Db    607  RSLDFTIPPVAGHNSCSIIVGVSGGGKIHVSPYGSKDVSLISAPIQPCELFNEVYCDTCT  666

Qy    316  ---------------APTTPKEPAPTT---PKEPAPTTKEPS-PTTPKEPAPTTTKS-  354
                           ||||||||| :      ||||| |    |||||||:
Db    667  AKYGAIHSGYQTSADFVTTTAKPTTTTTGAPGQPTTTTTGSPSKPTTTTTTKATTTTTI  726

Qy    355  ----APTTTKEPAPTTTKSAPTTPKEPSPTTKEPAPTTPKKPAPTTPKKPAPTTPKEPA  410
                ||||||||  : || :|  :  | :
Db    727  LNPIITTTTQKPTTTTTKVPGKPPIATTTTTLKPIVTTTTKATTTTTTTTVPTT-----  781

Qy    411  PTTPKEPAPTTTKKPAP-----------------------------TAPKEPAP----  435
            |||||||   ||||                                 ||||||
Db    782  TTTTKRDEMTTTTTPLPIDGIDEITPIPIEKMLDKYTRMIYDYNSGLLLLDSNDEPIPGSQ  841
```

```
;    FILING DATE: 07-AUG-1987
;    APPLICATION NUMBER: 933,945
;    FILING DATE: 24-NOV-1986
;    APPLICATION NUMBER: 650,128
;    FILING DATE: 13-SEP-1984
;SEQ ID NO:25:
    LENGTH: 744
5202236-25

Query Match
Best Local Similarity    7.7%;          Score 506.5; DB 6;   Length 744;
            Similarity  29.9%;   Pred. No. 1.9e-26;
Matches 239;   Conservative 78;   Mismatches 308;   Indels 175;   Gaps 50;

Qy   75 PKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKE 134
           |||:|||                              ::||
Db   24 PKMTYPPTYKPKPSYPP------------------------------TYKSKPTY----- 53

Qy  135 TQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKE----PASTTPKE 190
         |      ::|||  |: ||                :||
Db   54 ------------YPPTYK--AKPS-------------YPPTYKPKKTYPPTYKPKLTYPPTYKPKP 92

Qy  191 PTPTTIKSAPT-TPKEPAPTTTKSAPTTPKEPAPTTPKEPA---P 247
        |  |||  ||| ||:||  ::|||  ||
Db   93 SYPPTYKSKPTYKPKITYPPTYKAKPSYPPTYKPKKTYPPT-YKPKLTYPPTYKPKASYP 151

Qy  248 TTTKSAPTTPKEPAPTTPKKPAPT-TPKEPAPTT----PKEPTPTTPKEPAPTTKEP--- 299
           ||  ||||  ||         ||                :|   ||
Db  152 PTYKPKPSYP--PSYKTKKTYPPTYKPKLTYPPTYKPKPSYPPSYKPKKTYPPTYPPTYKPKLT 209

Qy  300 -APTTPKEPA--PTAPKKPA--PTTPKEPA--PTTTKEPS--PTTPKEPA 348
             ||||  |:    ||:       ||      ||
Db  210 YPPTYKAKPSYPPTYKAKPSYP--PTYKAKPSYPPTYKAKPSYPPTYKAKPS 269

Qy  349 -PTTTKSAPTTPKEPAPTTTKSAPTTPKEPSPTTPKEPA-----PTTPKKP 401
           ||   ||||  :|   ||   ::|||
Db  270 YPSTYKAKPSYT---YPSTYKAKPTYP----PTYKAKPSYPPTYKAKPTYKAKPTYKAKP 322

Qy  402 A--PTTPKEPA--PTTPKEPA--PTTTKKPA--PTAPKEPAPTTPKETAPTTPKKLTPTT 453
              ||||      ||||      ||    |
Db  323 SYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKVKPTYKAKPTYPS 382

Qy  454 PEKLAPTTPE--KPAPTTPEELAPTTPEPT--PTTPEEPA-PTTPKA--AAPNT-----P 502
                  ||:    |||:        ||      ||
Db  383 TYKAKPSYPPTYKAKPSYP-----PTYKAKPSYPPTYKAKPTYPSTYKAKP 438

Qy  503 KEPAPTTPKEPAPTT----PKEPAPTTPKETAPTT--PKGTAPTTLKEPAPTTPKKPAPK 556
          |   ||   |||       :|   |||
Db  439 SYPPTYKPKISYPPTYKAKPSYPSTYKAKSSYPPTYKAKPSYPSTYKAKP 498

Qy  557 ELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKGTAPTTLKEPA- 615
            :|||    :|   |||          ||       |
Db  499 SYPPTYKAKPTYPSTYKAKPTYK----ARPSYP-----PTYKAKPS------YPPTYKAKPSY 546

Qy  616 -PTTPKKPAPKELAPTTTKGPT--STTSDKPAPTTPKETAPTTPKKPAPKKPAPK- 670
            ||   |||     ||   |     |   |||
Db  547 PPTYKAKPTYK-AKPTYKAKPTYPSTYKAKPSYSPTYKAKPSYP----PTYKAKPSYPPT 601

Qy  671 ---TPETPPPTTSEVSTPTTTKE---PTTIHKSPD-----ESTPELSAEPT-PKALENS 717
               |   |||          |      ||
Db  602 YKAKPSYPPTYKAKPSYPPTYKAKPTYKAKPTYPSTYKAK 661

Qy  718 PKEPGVPTTKTPAA-------TKPEMTTTAKDKTTERDLRTTPETTTAAP------KMT 763
           |   |||               |
Db  662 PSYP--PTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKPTYPSTYKAKPSYPPTHKAKPT 719

Qy  764 KETATTTEKTTESKITATTT 783
               |      ::||
Db  720 YKAKPTYPSTYKAKPTYPST 739

RESULT 4
US-09-103-429A-3
```

```
Qy  436 --------------TTPKETA-PTTP-------------KKL 449
                        |   |   ||||
Db  842 AGQIADTSNLFPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSVSYLAAKNL 901

Qy  450 T----------------PTTPEKLAPTTPE------------------ 463
        |                ||||     |
Db  902 TVDTDETYGLPIDTLTGYPLPDPVSLIPFNPETGELFDPISDEIMNGTIAGIVSGISASES 961

Qy  464 ----KPAPTTPEE---------LAPTT---------PEEPTPTTP--EEPAPTTPKAAA 498
            |    ||            ||            |||    |
Db  962 LLSQKSAPIDPATNMVVGEFGGLLNPATGVMIPGSLGPSEQTPFSPEIEDGGIIPPEVAA 1021

Qy  499 PNTPKEPAPTTPKEPAPTTPKE------------------------------PA 522
         |  |   ||   |||                               |:
Db 1022 ANADKFKLSIPPSPVPESIPEKDQKIDSISELMYDIESGRLIGQVSKRPIPGSIAGDLNPI 1081

Qy  523 PTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKEPTSTTSDKPAPTTPKGT 582
         |||   |||  ||  |||                             |   |||  ||
Db 1082 MKTPTQDSVTGKPIDPTT---GLPFNP--PTGHLINPTNNNTMDSSFAGAYKYAVSNGI 1136

Qy  583 APTT------PKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP-----------KKPAPK- 625
         |||       |   ||                     ||                :
Db 1137 KTDNVYGLPVDEITGLPKDPVSDIPFNSTTGELVDPSTGKFINNYTAGIVSGKRGLPPIE 1196

Qy  626 ----ELAPTTTKGP----------TSTT--------SDKPAPTTPKETAPTTPKEPA--- 660
            |  |||   |         |  |         |   |||        |
Db 1197 DENGNLFDPSTKLPIDGNNQLVNPETNSTVSGSTSGSTKPKPGIPVNGGGVVPDEEAKDQ 1256

Qy  661 -----------PTTPKKPAPTTPETPPPTTSEVSTPTTTKE-PTTIHKS---PDESTPEL 705
                    |||   |||   |   |||       ||
Db 1257 ADKGKDGLIVPPTNSINKDPVTNTQYSNTTGNIINPETGKVIPGSLPGSLNYPSFNTPQQ 1316

Qy  706 SAEPTPKALE----NSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKM 762
          | |||       |  |   |||      |        :   |
Db 1317 TDEITGKPVDPVTGLPYDPSTGEIIDPATKLPIPGSVAGDEILTEVLNITTDEVTGLP-I 1375

Qy  763 TKETATTTEKTTESKITATTTQVTSTTQDTTPFKITTLKTTTLAPKVTTTKKITTTTE- 821
          |   |||      |   |||     :|   |   ||
Db 1376 DLETGLPRDP----------VSGLPQLPNGTLVD------------PSNKKPIPGSHS 1411

Qy  822 --IMNKPEETAKPKDRATNSKATTPKQKP-TKAP--KKPTSTKKPKTMPRVRRKT--- 873
            |    |   |        |||    |   ||    ||
Db 1412 GFINGTSGEQSHEKDPSTG------KPLDPNTGLPFDEDSGSLINPETGDKLQGSHSGTF 1465

Qy  874 TPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPK-SEDA---------- 922
         |   |   |   |         :| |||   |   |:   |
Db 1466 MPVPGKPQGENGGIMTPEQILEAL-----NKLPTSNEVNISPRPSSDAVPDRPTNTWW 1518

Qy  923 ---------GGAEGETPHMLL------RPHVFMPEVTPDMDYLPRVPNQGIII 960
                  |   |   :|:        |   |   ||
Db 1519 NKISGQTFQVDGKKTIPGSAASVIHTALGTPTQTDPTTGLP--SDPSTGLPFIPGFNVLV 1576

Qy  961 NPMLSD------------ETNIC--------------NGKPVDGLTTL---RNGTLV-AF 990
          |   |            |:  |                | |   ||        |
Db 1577 DPQTGEQIKGSVPVPVSLIYKEKNIVTEAAYGLPVDPKTGFPIDPISYLPFAKNGELIDPI 1636

Qy  991 RGHYF 995
         |  ||
Db 1637 SGKYF 1641

RESULT 3
5202236-25
;Patent No. 5202236
APPLICANT: MAUGH, KATHY J.;ANDERSON, DAVID M.;STRAUSBERG,
SUSAN L.;MCCANDLISS, RUSS;WEI, TENA;FILPULA, DAVID
TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
PROTEIN
NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/528,762
FILING DATE: 25-MAY-1990
APPLICATION NUMBER: 82,456
```

```
; Sequence 3, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
;; APPLICANT: Granados, Robert R
;; APPLICANT: Wang, Ping
;; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
;; TITLE OF INVENTION: cDNA and Related Products and Methods
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;;; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
;;; STREET: 118 No. 6187558th Tioga
;;; CITY: Ithaca
;;; STATE: NY
;;; COUNTRY: USA
;;; ZIP: 14850
;; COMPUTER READABLE FORM:
;;; MEDIUM TYPE: Floppy disk
;;; COMPUTER: IBM PC compatible
;;; OPERATING SYSTEM: PC-DOS/MS-DOS
;;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;;; APPLICATION NUMBER: US/09/103,429A
;;; FILING DATE: 24-JUN-1998
;;; CLASSIFICATION: 800
;; ATTORNEY/AGENT INFORMATION:
;;; NAME: Michaels, Christopher A
;;; REGISTRATION NUMBER: 34,390
;;; REFERENCE/DOCKET NUMBER: BTI-39
;; TELECOMMUNICATION INFORMATION:
;;; TELEPHONE: (607) 256-2000
;;; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;;; LENGTH: 786 amino acids
;;; TYPE: amino acid
;;; STRANDEDNESS: single
;;; TOPOLOGY: linear
;;; MOLECULE TYPE: protein
;;; HYPOTHETICAL: NO
;;; ANTI-SENSE: NO
;;; ORIGINAL SOURCE:
;;;; ORGANISM: Trichoplusia ni
;;;; TISSUE TYPE: peritrophic membrane
US-09-103-429A-3

Query Match     7.6%;   Score 498.5;   DB 4;   Length 786;
Best Local Similarity   29.1%;   Pred. No. 7.1e-26;
Matches  200;   Conservative   25;   Mismatches 207;   Indels 255;   Gaps  32;

Qy  156  PTPKAETTTKGPALTTPKEPTPTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTKSAP 215
              |||  | ||: ||: ||                                :|| ||
Db   93  PGPPAE-TTQAPATT----QAPTTT-------QAPTTTTQ--ATTTTQAP 135

Qy  216  TTPKEPAPTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 275
              ||: ||||| : |||: ||||| : |||| : |||| | ||||
Db  136  TTTQ--APTTQ--ATTT--QAPTTTQ-APTTTQ--APTTTQ--APTTTQ- 183

Qy  276  PAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA- 332
              ||| ||: : | ||   |||| : || ||||  ||
Db  184  -APTTTQAPTTTTQAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 241

Qy  333  ---------------------------------------------PTTTKEPSPTTPK- 345
                                                            : ||
Db  242  ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 301

Qy  346  ----EPAPTT---------------------------------------------------- 352
              ||| :
Db  302  DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 361

Qy  353  ------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 385
                                :|  |      ||||| :  ||||| :
Db  362  CGAGTHFSFLQQCDHIELVGCTLPGGSEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 420

Qy  386  APTT------PKEP------------------------------APTTPKKPAPTT--- 405
              |: |     ||                               :||     :||
Db  421  ADFSIDHLLPHESDCGYLQCVHGQTIARPCGNLHFSPATQSCESPVTAGCQVFECDSD 480

Qy  406  ---PKEPAPTTPKEPAPTTPKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 462
              ||| |||||: | ||||  ||  : |||  :|||
Db  481  NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT--AIPT-- 533

Qy  463  EKPAPTTPEELAPTTPEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 520
              |||| |||| |||  ||| |  || |||  |||    :||
Db  534  --PAPTAAPTAAPTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT--SAPTAAPT 587

Qy  521  PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTPKEPTSTTSDKPAPTTPK 580
              ||||| | | | |  |    |  |||      | : :||
Db  588  AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNT------------VTVPP 632

Qy  581  GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 640
              |||  |||  :||:  ||| :|
Db  633  TAAPTT-AAPAPNTTVTVPPTAAPTTAAPPTVAH--------------APNTTAAPVTTTS 677

Qy  641  DKPAPTTPKETAPTTPKEPAPTTPKKP 667
              |: |   |   || | |  |
Db  678  ---APATTPEDDDIDP--PLPNDPINP 699

RESULT 5
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;; APPLICANT: PETERSEN, CAROLYN
;; APPLICANT: LEECH, JAMES
;; APPLICANT: NELSON, RICHARD, C.
;; APPLICANT: GUT, JIRI
;; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;; TITLE OF INVENTION: INFECTIONS
;; FILE REFERENCE: 480.19-4(HV)
;; CURRENT APPLICATION NUMBER: US/08/700,651B
;; CURRENT FILING DATE: 1997-08-14
;; EARLIER APPLICATION NUMBER: 08/415,751
;; EARLIER FILING DATE: 1995-04-03
;; NUMBER OF SEQ ID NOS: 15
;; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
;; LENGTH: 1721
;; TYPE: PRT
;; ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match     7.4%;   Score 489;   DB 3;   Length 1721;
Best Local Similarity   22.1%;   Pred. No. 7.8e-25;
Matches  284;   Conservative   82;   Mismatches 476;   Indels 442;   Gaps  41;

Qy   29  KKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPS 88
              :  |
Db  116  RSNETK---TTEPSANTYAGVYRSN-ETKTTEPSANT---NFLLVDPKI----------N 158

Qy   89  LPPNSDTSKE------------------TSLTVNKETVETKETTTNKQTSTDGKEKTTSAK 133
              :|: :                     :||| : :  |||| ::|
Db  159  APCNSENSFEQGGIFDMGSKVVIPYTKCVGVKHTTTTTTTTTTTTTTTT-------- 208

Qy  134  ETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTP 193
                                                       |||  ||||||  |||
Db  209  --------------------------------------TTTTTTTTTTTTTTTTTTTTTT 240

Qy  194  TTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKSA 253
              || :|| ||||||    |||         ||
Db  241  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300

Qy  254  PTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTKEPAPTAPK 313
```

US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PETERS, VERNY, JONES & BIKSA
;   STREET: 385 Sherman Avenue, Suite 6
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94306-1840
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/928,361B
;   FILING DATE: 12-SEP-1997
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/026,062
;   FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: Verny, Hana
;   REGISTRATION NUMBER: 30,518
;   REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650-324-1677
;   TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1721 amino acid
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-928-361B-6

Query Match   7.4%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 7,7e-25;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

Qy   29 KKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPS 88
          ||| :| ||| ||: || | : :| | :| : | : ||| ||||
Db   116 RSNETK---TTEPSANTYAGVYRSN-ETKTTEPSANT---NFLLVDPKI---------N 158

Qy   89 LPPNSDTSKE------------------------------TSLTVNKETTVETKETTTNKQTSTDGKEKTTSAK 133
          : :| ||| : || | ||| || |:|
Db   159 APCNSENSFEQGQIFDMGSKVYIPYTKCVGVKHTTTTTTTTTTTTTTTTTTTT------ 208

Qy   134 ETQSIEKTSAKDLAPTSKVLAKPTKAETTKGPALTTPKEPTPTTPKEPASTTPKEPTP 193
          :| | : | | : : | ||| | || | ||||||||
Db   209 -------------------------------------TTTTTTTTTTTTTTTTTTTTT 240

Qy   194 TTIKSAPTTPKEPAPTTKSAPTTPKEPAPTTKEPAPTTPKEPAPTTKEPAPTTTKSA 253
          || :| |||||| : ||| |||| || | : :: || |||| |||||||||||
Db   241 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300

Qy   254 PTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTAPK 313
          :||:| || ||| | | :| | :: ::| |||| |||||||:
Db   301 TTTTKKPTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTT 350

Qy   314 KPAPTTPKEPAPTTPKEPSPTTPKEPAPTTTKSAPTTKEPAPTTTKSAPTT 373
          || || ||| :| || || |||| ||| || ||||||||||
Db   351 ---TTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTATTTTT 406

RESULT 6

APPLICANT: Hedstrom, Richard
APPLICANT: Khusmith, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: immunogen and gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
    ADDRESSEE: A. David Spevack
    STREET: NMRDC Building 1 T-12 National Naval
    STREET: Medical Center
    CITY: Bethesda
    STATE: MD
    COUNTRY: USA
    ZIP: 20814-5044
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/07/638,431
    FILING DATE: 19910110
    CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
    NAME: Spevack, Avrom D.
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (301) 295-6759
    TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 826 amino acids
        TYPE: AMINO ACID
        TOPOLOGY: linear
    MOLECULE TYPE: protein

US-07-638-431-2

```
Query Match        7.4%;   Score 488.5;  DB 1;  Length 826;
Best Local Similarity  26.4%;   Pred. No. 3.5e-25;
Matches  153;  Conservative  54;  Mismatches 223;  Indels 149;  Gaps 34;

Qy  192 TPTTIKSAPTTPKEPA-PTTTKSAPTTPKEPA-PTTTKEP----APTTPKEP-APTTKE 244
Db  273 TPCKVRDCPQIPIPPVIPNKIPEKPSNPEEPVNPDPNDPNNPNNPNNPNNPNNPNN 332

Qy  245 PAPTTTKSAPTTPKEP-APTTPKKP-APTTPKEP-APTTPKEP-TPTTPKEPAPTTPKEPA 300
Db  333 PNPNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNDPSNPNNHPK 392

Qy  301 PTTPKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTT 358
Db  393 RRNPKRRNPNKPKPNKPNKPNPNEPSNPNKPNPN-----EPSNPNKPNP-----NEPSN 443

Qy  359 TKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPA-PTTPKKP-APTTPKEPAPTTPKE 416
Db  444 PNKPNP---NEPSNPNKPNPN-----EPLNPNEPSNPNEPSNPNAPSNPNE--PSNPNE 492

Qy  417 PAPTTTKKPAPTAPKEP-APTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAP 475
Db  493 PS-------NPNEPSNPEPSNPNE--PSNPKK----------PSNPNE--P 523

Qy  476 TTPEEPTPTTPEEPAPTTPKAAAPNTPKEP-APTTPKEPAPTTPKEPAPTTPKETAPTTP 534
Db  524 SNPNE-PLNPNEP------SNPNEPSNPNEPSNPNEE--PSNPKE--PSNPNE------ 564

Qy  535 KGTAPTTLKEPAPTTPKKPAPKELAPTTPKGTPSTTSDKPAPTTPKGTAPTTPKETAPTT 594
Db  565 ----------PSNPEEPNPEE--PSNPKEP----SNPEEPINPEELNPKEPSNPEESN 606

Qy  595 PKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPT 654
Db  607 PKEPI----------NPEESNPKEP----INPEDENENPLIIQDEPIEPRNDSNVIPI 649
```

```
Qy  374 PKE-----------------------------------------------PSPTT------- 381
Db  407 TSETESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRIENGMAFTMIPNDDTHVRFRFK 466

Qy  382 ---------------------------TKEP--------------------------APTTPKE 392
Db  467 VKDVGNTISVRCRKGAGKLEFPDRSLDFTIPPVAGHNSCSIIVGVSCDGKIHVSPYGSKD 526

Qy  393 PAPTTPKKPAPTTPKE------------------------PAPTTPKKPAPTAPKEPA 434
Db  527 VSLIS----APIQPSELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTTAKPTTT--TGA 580

Qy  435 PTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTP 494
Db  581 PGQPTTTTTGSPSKPTTTTTKATTTT---TTLNPIITTTTTQKPTTTT-----TT- 627

Qy  495 KAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPA 554
Db  628 --KVPGKPPIATTTTTLKPIVTTTTTKATTTTTTTVPTT----TTTTKRDEMTTTTTPL 680

Qy  555 PK---ELAPTTTKEPTS---------------------TTSDKPAPTTPKG-TAPTTPKEPAPTT 594
Db  681 PDIGDIEITPIPIEKMLDKYTRMIYDYNSGLLLDSNDEPIPGSQAGQIADTSNLFPVQTH 740

Qy  595 PKEPAPTTPKGTAPTTLKE--PAPTTPKKPAPKELAPTTPKGTPSTTSDKPAPTTPKET 651
Db  741 KSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSVSYLAAKNLTVDTDETYG--LPIDT 798

Qy  652 APTTPKEPAPTTPKKPAPTTPETPPTTSEVSTPT-----------TTKEPTTIHKS------ 697
Db  799 LTGYPLDPVSLIPFN--PETGELFDPISDEIMNGTIAGIVSGISASESLLSQKSALIDPA 856

Qy  698 PDEST-------------------PELSAEPTPKALENSPK 719
Db  857 TNMVVGEFGGLLNPATGVMIPGFLGPSEQTQFSPEIEDGGIIPPEVAAANADKFKLSIP- 915

Qy  720 EPGVPTTKTPAATKPEMTTTAKDKTTER-----------------DLRTTPETTTAAPKM 762
Db  916 -PSVP------ESIPE-----KDQKIDSISELMYDIESGRLIGQVSKRPIPGSIAGDLNP 963

Qy  763 TKETATTTEKTTESKITATT-TQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTE 821
Db  964 IMKTPTQTDSVTGKPIDPTTGLPFNPPTGHLINPTNNNTMDSSFAGAYKYAVSNGIKTDN 1023

Qy  822 IMNKP-------EETAKPKD---------RATNSKATTPKPQKPTKAPKKPTSTKKPKTMP------ 866
Db 1024 VYGLPVGEITGLPKDPGSDIFPNSTTGELVDPSTGKPINNSTAGIVSGKPGLPPIEDENG 1083

Qy  867 --------------------RVRKPKTTPTPRKMTS-TMPE-------- 886
Db 1084 NLFDPSTNLPIDGNNQLVNPETNSTVSGSTSGTTKPKPGIPVNGGGVVPDEEAKDQADKG 1143

Qy  887 -----LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMP 941
Db 1144 KDGLIVPPTNSINKDPVTNTQYSNTTGNI----INP--ETGKVIPGSLPGSLNYPSFNTP 1197

Qy  942 EVTPD-------MDYLPRVPNQGIIINP----------MLSDETNICNGK----- 974
Db 1198 QQTDEITGKPVDTVTGLPYDPSTGEIIDPATKLPIPGSVAGDEILTEVLNITTDEVTGLP 1257

Qy  975 -------PVDGLTTLRNGTLV 988
Db 1258 IDLETGLPRDPVSGLPNGTLV 1281
```

```
Qy  655 TPKEP------APTTPKKPAPTTPETPPPTSEVSTPTTKEPTTIHKSPDESTPELSAE 708
            |:   |:.:.|| :| .|||  |:  .: || |  ||     |:||
Db  650 LPIIPQKGNNIPSNLPENPSDSEVEYPRPNDNGENSNNTMKSKKNI---PNEPIPSPGDN 706

Qy  709 PTPKALENSPK------------------------------EPGVPTTK 727
            || :|.|                              || .:|:|
Db  707 PYKGHEERIPKPHRSNDDYVYDNNVNKNKDEPEIPNNE 745
```

RESULT 8
PCT-US92-00018
; Sequence 2, Application PC/TUS9200018
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khusmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: immunogen and gene encoding
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMRDC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00018
; FILING DATE: 19920103
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-00018-2

Query Match   7.4%; Score 488.5; DB 5; Length 826;
Best Local Similarity 26.4%; Pred. No. 3.5e-25;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;
```

```
Qy  192 TPTTIKSAPTTPKEPA-PTTTKSAPTTPKEPA-PTTTKEP----APTTPKEP-APTTTKE 244
            || :: || |  | | :: |  | | | |  |    |    || |||| :| ||
Db  273 TPCKVRDCPQIPIPPVIPNKIPEKPSNPEPVNPNDPNDPNNPNPNNPNPNNPNN 332

Qy  245 PAPTTTKSAPTTPKEP-APTTPKKP-APTTPKEP-APTTPKEP-TPTTPKEPAPTTKEPA 300
            : | | |  |  |   |  |  |  |  | |   | |  | |:  | |
Db  333 PNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNDPSNPNNHPK 392

Qy  301 PTTPKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTT 358
            : |  | |    | | |  |  |  |  |  | :|  | :| ||| |  |  |
Db  393 RRNPKRRNPKPKPNKPNPNKPNPNEPSNPNKPNPN----EPSNPNKPNP----NEPSN 443

Qy  359 TKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKKPA-PTTPKKP-APTTPKEPAPTTPKE 416
            | |  |  :  |  |  | || |  |  | | |  | | |   |  | | :| :|||:
Db  444 PNKPNP----NEPSNPNKPNPN-----EPLNPNEPSNPNAPSNPNE--PSNPNE 492

Qy  417 PAPTTTKKPAPTAPKEP-APTTPKETAPTTPKLTPTTPEKLAPTTPEKPAPTTPEELAP 475
```

RESULT 9
5202236-13
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO:13:
; LENGTH: 652
5202236-13

```
Query Match     7.3%;  Score 476.5;  DB 6;  Length 652;
Best Local Similarity  29.7%;  Pred. No. 1.7e-24;
Matches  221;  Conservative  71;  Mismatches  284;  Indels 169;  Gaps 45;
```

```
Qy  75 PKITTAKPINPRPSLPPNSDTSKETSLIVNKETTVETKETTTNKQTSTDGKEKTTSAKE 134
           |:|  |  .:|: |                :.:| :|   | ||  ||
Db  11 PKMTYPPTYKPKPSYPP---------------TYKSKPTY----------KPKIT----- 40

Qy  135 TQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKEPTPTPKE----PASTTPKE 190
           | :  |  || |    |  ||   |  ::  | ||  : |||:| ||
Db  41 ------------YPPTYK--AKPS---------YPPTYKPKKTYPPTYKPKLTYPPTYKPKP 79

Qy  191 PTPTTIKSAPT-TPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPA--P 247
           | ||| |  |  |  | |  :  |  |  | | |||:  |:|  |:| |  |
Db  80 SYPPTYKSKPTYKPKITYPPTYKAKPSYPPTYKPKKTYPPT--YKPKLTYPPTYKPKASYP 138

Qy  248 TTTTKSAPTTPKEPAPTTPKKPAPT-TPKEPAPTT----PKEPTPTTPKEPAPTTKEPAPT 302
           |   | |  |||:| || |:  ||  |||:  |||    || |||| ||:| :| ||
Db  139 PTYKPKPSYP--PSYKTKKTYPPTYKPKLTYPPTYKPKSYPPSYKPNKTYPPTYKPKLT 196

Qy  303 TPKEPAPTAPKKPA--PTTPKEPA--PTTPKEPAPTTTKEPSPTTP----KEPAPTTTK 353
           |||:| || |:|   |||:| |   |||:| |||| ||:|| ||      |:| |||:|
Db  197 YP----PTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPSLIKAKPSYPPTYK 252

Qy  354 SAPTTTKEPA-------PTTTKSAPTTPKEPSPTTTKEPA--PTTPKEPA----PTTPKK 400
           || || |:|        |||:|  || |:| || |:|:|   |||:| |    |||:|:
Db  253 AKPTYKAKPTYKAKPTYPSTYKAKPTYP---PTYKAKPSYPPTYKAKPTYKAKPTYKAK 308
```

```
Qy  270  PTTPKEPAPTTP---KEPTPTTPKEPA------------------------  308
Db  187  P--PGAPGPAGPPGSRDPDGPPGAPGPAGPPGSRDPGPPGAHGPAGPKGAHG  244
Qy  309  --------PTAPK-KPAPTTP---KEPAPTTPKEPA-PTTTKEPSPTTPKEPAPTTTKS  354
Db  245  PAGPKGAHGPAGPKGAPGPAGPPGSRDPGPPGAPGPAGPPGSRDPGP--PGAPGP----A  298
Qy  355  APTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTP---KKPAPTTPKEPAP  411
Db  299  GPPGSRDPGP------PGAPGPAGPPGSRDPGP--PGAPGPAGPPGSRDPGP--PGAPGP  348
Qy  412  TTP---KEPAPTTKKPA------PTTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTT  461
Db  349  AGPPGSRDPDGPGPPGAPGPAGPPGSRDPGP----PGAPGPAGPPG  404
Qy  462  PEKPAPT-TPEELAPTTPEEPTPTTPEEPAPTTPKAAA-PNTPKEPAPTTP--KEPAPT  516
Db  405  SRDPGPPGAPGPAGPPGSRDPGP--PGAPGPAGPPGSRDPGPPGAPGPAGPPGSRDPGPP  462
Qy  517  TPKEPA------PTTPKET-APTTPKGT--------APTTLKEPAPTTPKKP-APKELA  559
Db  463  GAHGPAGPKGAHGPAGPKGAHGPAGPKGAPGPAGPPGSRDPGPPGAPGPAG  522
Qy  560  PTTTKEPTSTTSDKPAPTTPKGTA-PTTPKEPAPTTP---KEPAPTTPKGTA-PTTLKEP  614
Db  523  PPGSRDPGPPGA--PGPAGPPGSRDPGPPGAPGPAGPPGSRDPGPPGAPGPGPGSRDP  580
Qy  615  APTTPKKPAP------KELAPTTTKGPTST--TSDKPAPTTPKETAPTTPKEPAPTTPKK  666
Db  581  GP--PGAPGPAGPPGSRDPGPPGAPGPAGPPGSRDPGP-PGAPGPAGPPGSRDPGP--PGA  636
Qy  667  PAPTTP----ETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALEN------S  717
Db  637  PGPAGPPGSRDPGPPGAPGPAGPPGSRDPGP-PGAPGPAGPPGSRDPGPPGAPGPAGPPG  695
Qy  718  PKEPGVPTKTPAATK  733
Db  696  SRDPGPPGAHGPAGPK  711
```

RESULT 11
US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:

```
Qy  401  PA--PTTPKEPA--PTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEK  456
Db  309  PSYPPTTYKAKPSYPPTTYKAKPSYP----PTYKAKPSYPPTTYKVKPT--YK  362
Qy  457  LAPTTPE--KPAPTTPEELAPTTPEEPT--PTTPEEPA-PTTPKAAAPNTPKEPAPTTPK  511
Db  363  AKPTYPSTTYKAKPSYP----PTYKAKPSYPPTTYK-AKPSYP----PTYKA  413
Qy  512  EPA-PTT----PKEPAPTTPKETAPTT--PKGTAPTTLKEPAPTTPKKPAPKELAPTTTK  564
Db  414  KPTYPSTTYKAKPSYPPTYKPKISYPPTTYKAKPSYPSTTYKAKSSYPPTTYKAKPSYPPTTYKA  473
Qy  565  EPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPA--PTTPKKP  622
Db  474  KPTYPSTTYKAKPTYK--AKPSYP----PTYKAKPS------YPPTYKAKPSYPPTTYKAKP  521
Qy  623  APKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPA--PTTPKKPA-PTT----PETP  675
Db  522  TYK-AKPTYKAKPSYPPTTYKAKPSYP----PTYKAKPSYPPTTYKAKPSYPSTTYKAKPSYP  576
Qy  676  PPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPE  735
Db  577  PTYKAKPSYPPT-------YKAKPSYPPTYKAKPTYKA-----KPTYPST---YKAKPS  620
Qy  736  MTTTAKDKTTERDLRTTPETTTAAP  760
Db  621  YPPTHKAKPTYKAKPSYPPTHKAKP  645
```

RESULT 10
US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-5

Query Match 6.9%; Score 452; DB 4; Length 960;
Best Local Similarity 25.6%; Pred. No. 1.2e-22;
Matches 173; Conservative 60; Mismatches 309; Indels 134; Gaps 41;

```
Qy  156  PTPKAETTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAP--TTTKS  213
Db   72  PGPAGPPGSRDPG--PPGAPGPAGP--PGSRDPGPPGAPG-PAGPPGSRDPGPPGAPGPA  126
Qy  214  APTTPKEPAPTTTKEPA-PTTPKEPAPTTTKKEPAPTTTKSAPTTPKEPAPTTP---KKPA  269
Db  127  GPPGSRDPGPPGAPGPAGPPGSRDPGPPGAPGPAGPPGSRDPGPPGAPGPAGPPGSRDPG  186
```

APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1727 constitute a repeated region wherein the repeat i
20 amino acids, 17 of which are fixed. The number of such
repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
21 amino acid precursor sequence."
US-08-479-537A-5

Query Match   6.5%;
Best Local Similarity 23.7%;   Score 424.5; DB 2; Length 1867;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

```
Qy   99 TSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTTSAKD-LAPTSKVLAKPT 157
Db   16 TVLTV-------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS      65
Qy  158 P-KAETTKGP--ALTTPKEP------------TPTTPKEPATTPKEPTPTTIKSAP    200
Db   66 PGSGSSTTQGODVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTPPAAHDVT---SAP 122
Qy  201 TTPKEPAPTTP--------APTTTKEPAPTTPKEPAPTT                     241
Db  123 --DNKPAPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS                    180
Qy  242 TKE-PAPTTT------KSAPTTPKEP---KPAPTTPKEPAPTTPKEPTPTT         288
Db  181 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS                    240
Qy  289 PKE--PAPTTKEP------APTTPKEPAPTTPKEPAPTT                     335
Db  241 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS                    300
Qy  336 TKEPSPTTPKEPAPTT----TKSAPTTPKEPSPTTTK----E                 384
Db  301 APDXRP-XPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT                    359
Qy  385 PAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPK---EPAPTTTKKPAPTAPKEPAPT  436
Db  360 SAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT                    419
Qy  437 TPKETAP---TTPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTTPEEPAPT  492
Db  420 SAPDXRPXPGSTAPXAHGVTS---APDXRPXPGSTAPXAHGVTSAPDXRPX----PGST 471
Qy  493 TPKA---AAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGT-APTTLKEPAP         547
Db  472 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRP-XPGSTAPXAHGVTSAPDXRPXPGS 530
Qy  548 TTP--------KKPAPKELAPTTTKEPTSTTSDKPAP--TTPKGTAPTTPKEPAPTTP  595
Db  531 TAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDXRPXPGSTAPXAHGVTS----APDXR 585
Qy  596 KEPAPTTPKG---TAPTTLKEPAPTTP--------KKPAPKELAPTTTKGPTSTTSD   641
Db  586 PXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDX 644
Qy  642 KPAP--TTPK----ETAPTTPKEPAPTTP------KKPAPTTPETPPPTTSEVSTP    685
Db  645 RPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS-TAPXAHGVTSAP  702
Qy  686 TTTKEPTT------IHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA-------  731
Db  703 DXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAP 762
Qy  732 -TKPEMTTTAKDK---TTERDLRTTPETTTAAPKMTKETATTTEKTTSKITATTTQVTS 787
Db  763 DXRPXPGSTAPXAHGVTSAPDXRPXPGST-APXAHGVTSAPDXRPXPGSTAPXAHGVTS  820
Qy  788 TTTQDTTPFKITTLKTTLAPKVTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQ  847
Db  821 APDXRPXPG-----STAPXAHGVT------SAPDXRPXPGSTAPXAHGVTSAPDXRPXP- 868
Qy  848 KPTKAPKKPTSTKKPKTMPRVRKPKTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT 907
Db  869 -GSTAPXAHGVTSAPDXRP-----XPGSTAPXAHGVTSA-PDXRPX-----PGST     911
Qy  908 -PNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD 966
Db  912 APXAHGVTSAPDXRPXPGSTAPXAH-----------GVTSAPDXRPXPGSTA---PXAHG 957
Qy  967 ETNICNGKPVDGLTT-LRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPID   1021
Db  958 VTSAPDXRPXPGSTAPXAHGVTSA-------------PDXRPXPGSTAPXAHGVTSAPD  1003
```

RESULT 12
US-09-083-116-5
; Sequence 5, Application US/09083116
; Patent No. 6203795
GENERAL INFORMATION:
APPLICANT:
APPLICANT: CHAMBON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

```
Query Match            6.5%;  Score 424.5;  DB 4;    Length 1867;
Best Local Similarity  23.7%;  Pred. No. 1.8e-20;
Matches   255;  Conservative   69;  Mismatches  511;  Indels  241;  Gaps  53;

Qy    99  TSLTVNKETTVETKETTTNKQTSTDGKEKTSAKETQSIEKTSAKD-LAPTSKVLAKPT  157
              |   |||        ::  ||:|||  :::::|:|:  ::: ::
Db    16  TVLTV----------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS  65

Qy   158  P-KAETTTKGP-ALTTPKEP----------------------TPTTPKEPASTPKEPTTTIKSAP  200
          | :||| ||  ||||||                        ||| |||
Db    66  PGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVT---SAP  122

Qy   201  TTPKEPAPTTT--------KSAPTTPKEP-----------------APTTTKEPAPTTPKEPAPTT  241
            |:  :||          |||                      ||
Db   123  --DNKPAPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  180

Qy   242  TKE--PAPTTT-----KSAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPTPTT  288
          ||:    |         |||
Db   181  APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  240

Qy   289  PKE--PAPTTKEP-------APTTPKEPAPTAPK----KPAPTTPKEPAPTTPKEPAPTT  335
          ||:   |  |          |||                |
Db   241  APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  300

Qy   336  TKEPSPTTPKEPAPTT-----TKSAPTTKEPAPTT-----TKSAPTTPKEPSPTTK----E  384
          ||: :||            |||:
Db   301  APDXRP-XPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT  359

Qy   385  PAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPK---EPAPTTTKPAPTAPKEPAPT  436
          ||                  |
Db   360  SAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT  419

Qy   437  TPKETAP---TTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPT  492
          ||:          :||
Db   420  SAPDXRPXPGSTAPXAHGVTS----APDXRPXPGSTAPXAHGVTSAPDXRPX----PGST  471

Qy   493  TPKA----AAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKGT-APTTLKEPAP  547
          ||:     : ||
Db   472  APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRP-XPGSTAPXAHGVTSAPDXRPXPGS  530

Qy   548  TTP--------KKPAPKELAPTTKEPTSTTSDKPAP--TTPKGTAPTTPKEPAPTTP  595
          ||:           ||
Db   531  TAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDXRPXPGSTAPXAHGVTS----APDXR  585

Qy   596  KEPAPTTPKG---TAPTTLKEPAPTTP----------KKPAPKELAPTTTKGPTSTTSD  641
              ||             |
Db   586  PXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDX  644

Qy   642  KPAP--TTPK---ETAPTTPKEPAPTTP--------KKPAPTTPETPPPTTSEVSTP  685
          :|:                ||
Db   645  RPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS--TAPXAHGVTSAP  702

Qy   686  TTTKEPTT------IHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA----  731
             |:||
Db   703  DXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAP  762

Qy   732  -TKPEMTTAKDK---TTERDLRTTPETTAAPKMTKETATTTEKTESKITATTTQVTS  787
              |   |||                 |
Db   763  DXRPXPGSTAPXAHGVTSAPDXRPXPGST-APXAHGVTSAPDXRPXPGSTAPXAHGVTSAP  820

Qy   788  TTTQDTTPFKITTLKTTTLAPKVTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQ  847
                   |::   |||       :::
Db   821  APDXRPXPG-----STAPXAHGVT-----SAPDXRPXPGSTAPXAHGVTSAPDXRPXP-  868

Qy   848  KPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT  907
              ||                     |
Db   869  -GSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSA-PDXRPX----------PGST  911

Qy   908  -PNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD  966
                 ||:           |                    :
Db   912  APXAHGVTSAPDXRPXPGSTAPXAH----------GVTSAPDXRPXPGSTA---PXAHG  957

Qy   967  ETNICNGKPVDGLTT-LRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPID  1021
```

```
Db    958  VTSAPDXRPXPGSTAPXAHGVTSA---------PDXRPXPGSTAPXAHGVTSAPD  1003
                :|  :: :|   ||   :|              |:|   |   |  ::| |

RESULT 13
US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Patent No. 5861381
GENERAL INFORMATION:
   APPLICANT: CHAMBON, Pierre
   APPLICANT: KIENY, Marie-Paule
   APPLICANT: LATHE, Richard
   APPLICANT: HAREUVENI, Mara
   TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
   TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
   NUMBER OF SEQUENCES: 5
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
      STREET: P.O. Box 1404
      CITY: Alexandria
      STATE: Virginia
      COUNTRY: United States
      ZIP: 22313-1404
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/479,537A
      FILING DATE: 07-JUN-1995
      CLASSIFICATION: 514
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: FR 90/13101
      FILING DATE: 23-OCT-1990
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: WO PCT/FR91/00835
      FILING DATE: 23-OCT-1991
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/039,320
      FILING DATE: 04-APR-1993
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/403,576
      FILING DATE: 14-MAR-1995
   ATTORNEY/AGENT INFORMATION:
      NAME: Teskin, Robin L.
      REGISTRATION NUMBER: 35,030
      REFERENCE/DOCKET NUMBER: 017753-025
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: (703) 836-6620
      TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 2035 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
   MOLECULE TYPE: peptide
   FEATURE:
      NAME/KEY: Peptide
      LOCATION: 128..1899
      OTHER INFORMATION: /note= "The amino acids spanning
      OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat i
      OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
      OTHER INFORMATION: repeats varies from 1 to 40."
   FEATURE:
      NAME/KEY: Peptide
      LOCATION: 134
      OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
      OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
      OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
   FEATURE:
      NAME/KEY: Peptide
```

```
      LOCATION: 144
      OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
      OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
      OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
   FEATURE:
      NAME/KEY: Peptide
      LOCATION: 147
      OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
      OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
      OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
   FEATURE:
      NAME/KEY: Peptide
      LOCATION: 1..21
      OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
      OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-2

Query Match             6.5%;    Score 424.5;  DB 2;   Length 2035;
Best Local Similarity  23.7%;    Pred. No. 2e-20;
Matches   255;  Conservative   69;  Mismatches  511;  Indels 241;  Gaps  53;

Qy    99  TSLIVNKETTVETKETTTNKQSTDGKEKTSAKETQSIEKTSAKD-LAPTSKVLAKPT  157
                                                                 |:|| :||| ||:| :|| ||:| ::|| ||   :|: |||
Db    16  TVLIV----------------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS  65

Qy   158  P-KAETTTKGP-ALTTTKEP----------------------TPTTPKEPASTTPKEPTPTTIKSAP  200
           | |:|| ||| :: ||||||                      ||  |||||| |||||| ||:||  ::|
Db    66  PGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVT---SAP  122

Qy   201  TTPKEPAPTT------------KSAPTTPKEP-----------------APTTKEPAPTTPKEPAPTT  241
              ::||| :|                |||   :|                    |||
Db   123  --DNKPAFGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  180

Qy   242  TKE--PAPTT-------KSAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPTPTT  288
               |||         :|    |||                 |:|
Db   181  APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  240

Qy   289  PKE--PAPTTKEP-----------APTTPKEPAPTAPK---KPAPTTPKEPAPTTPKEPAPTT  335
              ||:         :|       ||        |||
Db   241  APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  300

Qy   336  TKEPSPTTPKEPAPTT-----TKSAPTTKEPAPTT-----TKSAPTTPKEPSPTTTK----E  384
                 |: ||   ::|||            |||             |||
Db   301  APDXRP-XPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT  359

Qy   385  PAPTTPKEPAPTTPK----KPAPTTPKEPAPTTPK---EPAPTTTKKPAPTAPKEPAPT  436
             |:                |:|                :|
Db   360  SAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT  419

Qy   437  TPKETAP----TTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPT  492
             ||                    :  :|       ||
Db   420  SAPDXRPXPGSTAPXAHGVTS----APDXRPXPGSTAPXAHGVTSAPDXRPX----PGST  471

Qy   493  TPKA----AAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGT-APTTLKEPAP  547
              |:              ::|:     :|:  ||  :|:  |:|
Db   472  APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRP-XPGSTAPXAHGVTSAPDXRPXPGS  530

Qy   548  TTP-------KKPAPKELAPTTTKEPTSTTSDKPAP---TTPKGTAPTTPKEPAPTTP  595
              ||          |:                           :|          :: -APDXR
Db   531  TAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDXRPXPGSTAPXAHGVTS----APDXR  585

Qy   596  KEPAPTTPKG----TAPTTLKEPAPTTP------KKPAPKELAPTTTKGPTSTTSD  641
             |:             ||                    |:
Db   586  PXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDX  644

Qy   642  KPAP--TTPK----ETAPTTPKEPAPTTP------KKPAPTTPETPPPTTSEVSTP  685
             |                      ::|              :|
Db   645  RPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS--TAPXAHGVTSAP  702

Qy   686  TTTKEPTT-------IHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA-------  731
             ||:||                  ::|                         |
Db   703  DXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAP  762
```

```
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: peptide
;  FEATURE:
;  NAME/KEY: Peptide
;  LOCATION: 128..1899
;  OTHER INFORMATION: /note= "The amino acids spanning
;  OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repea
;  OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
;  OTHER INFORMATION: repeats varies from 1 to 40."
;  FEATURE:
;  NAME/KEY: Peptide
;  LOCATION: 134
;  OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
;  OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
;  OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
;  FEATURE:
;  NAME/KEY: Peptide
;  LOCATION: 144
;  OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
;  OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
;  OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
;  FEATURE:
;  NAME/KEY: Peptide
;  LOCATION: 147
;  OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
;  OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;  OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
;  FEATURE:
;  NAME/KEY: Peptide
;  LOCATION: 1..21
;  OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
;  OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match                   6.5%;    Score 424.5;  DB 4;   Length 2035;
Best Local Similarity   23.7%;    Pred. No. 2e-20;
Matches  255;  Conservative  69;  Mismatches  511;  Indels  241;  Gaps  53;

Qy      99 TSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKD-LAPTSKVLAKPT 157
              ||  : :.: ||| ||||:: :: :::.||::||:: ||:::  ||:::
Db      16 TVLTV-------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS 65

Qy     158 P-KAETTKGP--ALTTPKEP-------------------TPTTPKEPASTTPKEPTPTTIKSAP 200
            |  :||:|                        |||   |||| |||
Db      66 PGSGSSTTQGGDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVT---SAP 122

Qy     201 TTPKEPAPTTT---------KSAPTTPKEP------------APTTKEPAPTTPKEPAPTT 241
                                ::| :||                   |
Db     123 --DNKPAPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 180

Qy     242 TKE--PAPTTT------KSAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPTPTT 288
                                      |||                   | :
Db     181 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 240

Qy     289 PKE--PAPTTKEP--------APTTKEPAPTTPK----KPAPTTPKEPAPTTPKEPAPTT 335
                                  | |:                  |::|
Db     241 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 300

Qy     336 TKEPSPTTPKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTPKEPSPTTTK----E 384
                                                          :
Db     301 APDXRP-XPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT 359

Qy     385 PAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPK----EPAPTTKKPAPTAPKEPAPT 436
                  |                         :|                    ||
Db     360 SAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXRPXPGSTAPXAHGVT 419

Qy     437 TPKETAP----TTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPAPT 492
            |::|                      |: |         |||
Db     420 SAPDXRPXPGSTAPXAHGVTS----APDXRPXPGSTAPXAHGVTSAPDXRPX----PGST 471

Qy     493 TPKA----AAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGT-APTTLKEPAP 547
```

```
Qy     732 -TKPEMTTTAKDK---TTERDLRTTPETTAAPKMTKETATTTEKTTESKITATTTQVTS 787
              :|   :||       :|   :|:    :  :  :                |||
Db     763 DXRPXPGSTAPXAHGVTSAPDXRPXPGST--APXAHGVTSAPDXRPXPGSTAPXAHGVTS 820

Qy     788 TTTQDTTTPFKITTLKTTLAPKVTTTKKTITTTBIMNKPEETAKPKDRATNSKATTPKPQ 847
            ::  |:     :                                              |
Db     821 APDXRPXPG-----STAPXAHGVT------SAPDXRPXPGSTAPXAHGVTSAPDXRPXP- 868

Qy     848 KPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT 907
             |::  : :| :                |: :                          |
Db     869 -GSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSA-PDXRPX-----------PGST 911

Qy     908 -PNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD 966
              :  : :                |: :                        :
Db     912 APXAHGVTSAPDXRPXPGSTAPXAH--------GVTSAPDXRPXPGSTA---PXAHG 957

Qy     967 ETNICNGKPVDGLTT-LRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPID 1021
            |:  :                   |: :                    | :
Db     958 VTSAPDXRPXPGSTAPXAHGVTSA---------PDXRPXPGSTAPXAHGVTSAPD 1003
```

```
RESULT 14
US-09-083-116-2
; Sequence 2, Application US/09083116
; Patent No. 6203795
;  GENERAL INFORMATION:
;    APPLICANT: CHAMBON, Pierre
;    APPLICANT: KIENY, Marie-Paule
;    APPLICANT: LATHE, Richard
;    APPLICANT: HAREUVENI, Mara
;    TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES: 5
;    CORRESPONDENCE ADDRESS:
;    ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;    STREET: P.O. Box 1404
;    CITY: Alexandria
;    STATE: Virginia
;    COUNTRY: United States
;    ZIP: 22313-1404
;    COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/09/083,116
;    FILING DATE:
;    CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: 08/479,537
;    FILING DATE:
;    PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: WO PCT/FR91/00835
;    FILING DATE: 23-OCT-1991
;    PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 08/039,320
;    FILING DATE: 04-APR-1993
;    PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 08/403,576
;    FILING DATE: 14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;    NAME: Teskin, Robin L.
;    REGISTRATION NUMBER: 35,030
;    REFERENCE/DOCKET NUMBER: 017753-025
;    TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (703) 836-6620
;    TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 2035 amino acids
;    TYPE: amino acid
```

```
;    TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2476 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-276-967-2

Query Match              6.4%;  Score 419.5;  DB 2;  Length 2476;
Best Local Similarity   32.5%;  Pred. No. 5.5e-20;
Matches  163;  Conservative  43;  Mismatches  177;  Indels  119;  Gaps  31;

Qy  430 PKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEPTPTTPEEP 489
            | |  :|| |||  |||  | :||  | |||  :  |  |  :  :|
Db  313 PSETSVSTEKPVAPTE----KPTVPSEIYTIPTEKPMVHMEKPIVHT--EKPTVPT-EKP 365

Qy  490 APTTPKAAAPNTPKEPA-----PTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 545
           |  ||  :||   :  |  |||  | |||  | ||  | :||
Db  366 TIPTEKSTVPT--KKPTVFKEPTLPPE-GPTVPAE-RPTTPPEGPAVPPKG--PTVLTE- 418

Qy  546 APTTPKKPAPKELAPTTKEPTSTTSDKPAPTTPKGTAPT-TPKEPAPTTPKEPAPTTPK 604
            |   :| | |    | :|  ||| | |||   :||           :|| ||
Db  419 ------------WPTSHTEKSTVHTEKPILPTGKSTIPTEKPMVPTKRT------TTP- 458

Qy  605 GTAPTTLKEPAPTTPKKPAPKELAPT-TTKGPTSTT---SDKPAPTTPKETAPTTPKEPA 660
           | | | |   |  :| |   |   | ||   || |     | :|| | :||  |:|
Db  459 -TERTTIPAEKPTV--PIEKPMVPTERTTIPTERTTIPTEKPTVPTEKLTVPT--EKPI 512

Qy  661 PTTPKKPAPTTPETPPP-----TTSEVSTPT-----TTKEPTTIHKSPDESTPELSAEPT 710
            |  :| |  | |  ||     |:   | ||      |||    ||
Db  513 VPTEKPIVPTEKHTIPTEKLTVLTERTTTPTERTTIPTERTTIPTEKPTVPTEKPSVPT- 571

Qy  711 PKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTT 770
           :|| | |   :  || :  | |   |   |  | :|| | |   |  |  |:
Db  572 ------EEPTIPTEKLTVPT--ERTTTPTKRTTTPTIRTTTPTI------RTTTPT 613

Qy  771 EKTTESKITATTQVTSTTTQDTTPFKITTLKTTLAPKVTTTKKTITTTEIMNKPETA 830
            |||     :|| |  :||| |   |:||     :|||  | :|| |
Db  614 ERTT------TPTIRTTTPTERTT--IPTKKTT------VPTEKTIIPT------ERTI 652

Qy  831 KPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPT 890
           :|    |  || |     | |      | ||    ::|||  | :||  :|
Db  653 AP--------TTPQP-SPTLVPTQPAAVVMPST------SATTVTPTTIASCP---PN 693

Qy  891 SRIAEAMLQTTTRPNQTPNSKL 912
             :|     |||  | :|   :|
Db  694 AHFERCACPVSCQ-SPTPNCEL 714

Search completed:  April 26, 2002, 16:25:14
Job time: 454 sec
```

```
Db  472 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRP-XPGSTAPXAHGVTSAPDXRPXPGS 530
            :||  |     |     | |  |||
Qy  548 TTP--------KKPAPKELAPTTTKEPTSTTSDKPAP--TTPKGTAPTTPKEPAPTTP 595
            | |       :| | | |  | |  ||| :   | |   ||| | |||  ||
Db  531 TAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDXRPXPGSTAPXAHGVTS----APDXR 585

Qy  596 KEPAPTTPKG----TAPTTLKEPAPTTP--------KKPAPKELAPTTTKGPTSTTSD 641
            :|| || |       | | | |  |  |       :| | | |   | |   ||| :|
Db  586 PXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDX 644

Qy  642 KPAP--TTPK----ETAPTTPKEPAPTTP--------KKPAPTTPETPPPTTSEVSTP 685
            : |    ||     | | || |  |  |        :| |  | | |  |:  | |
Db  645 RPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS--TAPXAHGVTSAP 702

Qy  686 TTTKEPTT-------IHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA------- 731
            || : ||        | | |  | |  | | :|| | |   :  ||  |
Db  703 DXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAP 762

Qy  732 -TKPEMTTTAKDK---TTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTQVTS 787
            |   |   |       | :|| | |   |  |  |:    |||     :|| |
Db  763 DXRPXPGSTAPXAHGVTSAPDXRPXPGST-APXAHGVTSAPDXRPXPGSTAPXAHGVTS 820

Qy  788 TTTQDTTPFKITTLKTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQ 847
            :||| |   |:||     :|||  | :|| |    :|    |  || |
Db  821 APDXRPXPG-----STAPXAHGVT------SAPDXRPXPGSTAPXAHGVTSAPDXRPXP- 868

Qy  848 KPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTRPNQT 907
            :|    | |      | |      ::|||  | :||  :|     :|     ||  | :|
Db  869 -GSTAPXAHGVTSAPDXRP-XPGSTAPXAHGVTSA-PDXRPX-----------PGST 911

Qy  908 -PNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD 966
            :   |                  |
Db  912 APXAHGVTSAPDXRPXPGSTAPXAH--------GVTSAPDXRPXPGSTA----PXAHG 957

Qy  967 ETNICNGKPVDGLTT-LRNGTLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPID 1021
            :   |
Db  958 VTSAPDXRPXPGSTAPXAHGVTSA----------PDXRPXPGSTAPXAHGVTSAPD 1003
```

RESULT 15
US-08-276-967-2
; Sequence 2, Application US/08276967
; Patent No. 5851817
GENERAL INFORMATION:
   APPLICANT: Hardy, Daniel M.
   APPLICANT: Garbers, David L.
   TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
   TITLE OF INVENTION: Sperm
   NUMBER OF SEQUENCES: 8
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: Arnold, White & Durkee
      STREET: P. O. Box 4433
      CITY: Houston
      STATE: Texas
      COUNTRY: USA
      ZIP: 77210-4433
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/276,967
      FILING DATE: Submitted Herewith
      CLASSIFICATION: 530
   ATTORNEY/AGENT INFORMATION:
      NAME: Kitchell, Barbara S.
      REGISTRATION NUMBER: 33,928
      REFERENCE/DOCKET NUMBER: UTSD:418\KIT
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: 713-787-1400
      TELEFAX: 713-789-2679

aal.rai

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     April 26, 2002, 16:30:58 ; Search time 114.61 seconds
                                        (without alignments)
                                        905.906 Million cell updates/sec

Title:          AA2
Perfect score:  7276
Sequence:       1 MAWKTLPIYLLLLLSVFVIQ.........ARAITTRSGQTLSKVWYNCP 1363

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters:     219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      PIR_68:*
                1: pirl:*
                2: pir2:*
                3: pir3:*
                4: pir4:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	1077	14.8	3020	2	A43932	mucin 2 precursor,
2	950	13.1	1664	2	T18262	s-layer protein -
3	863.5	11.9	1489	2	T31108	cyst germination s
4	846	11.6	1274	2	T16251	hypothetical prote
5	812	11.2	2187	2	T30826	nascent polypeptid
6	809	11.2	1367	2	S48878	glucan 1,4-alpha-g
7	784	10.8	1188	2	S49915	extensin-like prot
8	695	10.8	1151	2	T18535	high molecular mas
9	684.5	9.4	1229	2	T25697	mucin 1 precursor,
10	672	9.2	1344	1	A35175	elastic titin - hu
11	666.5	9.2	7962	2	I38346	mucin MUC5B, trach
12	653	9.0	3570	2	T45025	nucleolar phosphop
13	643.5	8.8	990	2	I51618	hypothetical prote
14	636	8.7	3507	2	T34513	extensin class 1 p
15	633	8.7	489	2	T11622	hypothetical prote
16	632	8.7	761	2	C84672	protein UNC-89 - C
17	629	8.6	6642	2	T29757	hypothetical prote
18	627.5	8.6	971	2	T19431	hypothetical prote
19	622.5	8.6	839	2	F75518	hypothetical prote
20	607.5	8.3	801	2	T29018	hypothetical prote
21	607.5	8.3	924	2	S27923	gene LF3 protein -
22	605	8.3	379	2	S50125	larval glue protei
23	591.5	8.1	2232	2	T34434	hypothetical prote
24	568.5	7.8	1630	2	A53577	ascites sialoglyco
25	559.5	7.7	350	2	S22456	hydroxyproline-ric
26	559	7.7	875	2	S23760	polyphenolic adhes
27	556.5	7.6	856	2	T16543	hypothetical prote
28	551	7.6	620	2	S06733	hydroxyproline-ric
29	551	7.6	1087	1	QFMSH	neurofilament trip
30	545	7.5	873	2	A47283	calphotin - fruit
31	542	7.4	369	2	S20500	hydroxyproline-ric
32	540	7.4	1459	2	T32271	hypothetical prote
33	538.5	7.4	416	1	JU0465	extensin precursor
34	537	7.4	1072	2	A37221	neurofilament trip
35	532	7.3	756	2	T27642	hypothetical prote
36	530.5	7.3	813	2	S70795	vsaA protein precu
37	530.5	7.3	1162	2	JH0557	exo-alpha-sialidas
38	530	7.3	865	2	A47282	calcium-binding pr
39	522	7.2	328	2	JQ0985	hydroxyproline-ric
40	521.5	7.2	854	2	S02003	neurofilament trip
41	518	7.1	866	2	T45462	membrane glycoprot
42	518	7.1	1611	2	T38236	hypothetical prote
43	515.5	7.1	3534	2	T42567	tegument protein 2
44	513	7.1	1832	2	T31113	mucin-like glycopr
45	512.5	7.0	867	2	T45463	membrane glycoprot

ALIGNMENTS

```
RESULT      1
A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A;Reference number: A49963; MUID:94132002
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A;Reference number: A45106; MUID:93016075
A;Accession: A45106
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A;Reference number: A43932; MUID:91358717
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350,'L',1352-1411,'S',1413-1448,'P',1450-1503,'T',1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749; NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A;Reference number: A33532; MUID:89197956
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
```

J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A;Reference number: PQ0328; MUID:92198477
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
C;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C;Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match   14.8%; Score 1077; DB 2; Length 3020;
Best Local Similarity 27.2%; Pred. No. 1.4e-37;
Matches 373; Conservative 103; Mismatches 497; Indels 398; Gaps 50;

Qy   36 SFERGRECDC-------DAQCKKYDKC-------CPDYESFCAEVHNPTSPPSSKKAP 79
               : ||:||:   :| || || | || ||
Db   1081 SCDTGDCECFCSAVASYAQECTKEGACVFWRTPDLCP-----IFC-DYYNPPHECEWHYEP 1136

Qy   80 PPSGASQTIKSTT--------------KRSPK-PPNKKKTKKVIESEE----ITEEH- 117
               : ::: || |: | ||| | || :||: :
Db   1137 CGNRSFETCRTINGIHSNISVSYLEGCYPRCPKDRPIYEEDLKKCVTADKCGCYVEDTHY 1196

Qy   118 ---SVSENQESSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVDNKKNRTKKKPTPK 174
                  |:   :|:
Db   1197 PPGASVPTEETCKSCVCTNSSQVVCRPEEGK----------ILNQTQ------ 1233

Qy   175 PPVVDEAGSGLDNGDF--KVTTPDTISTTQH--NKVSTSPKITTAKPINPRPSLPPNSD 228
               : | :| |::   :|:| :|: |
Db   1234 ----DGAFCYWEICGPNGTVEKHFNICSITTRRSTLTTFTTITLPTTP----- 1277

Qy   229 TSKETSLTVNKETTVETKET--TTTNK-------QTSTDGKEKTTS--AKET 269
               : | || | ::||: |:| | |:|
Db   1278 -TSFTTTTTTTTTPTSSTVLSTTPKLCCLWSDWINEDHPSSGSDDGDREPFDGVCGAP 1333

Qy   270 QSIEKTSAKD-------------------------- 279
               : ||
Db   1334 EDIECRSVKDPHLSLEQHGQKVQCDVSVGFICKNEDQFGNGPFGLCYDYKIRVNCCWPMD 1393

Qy   280 -LAPTSKVLAKPTSKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAP-TT 336
               : :||: | |:||:|||| | || ||||||||||||||
Db   1394 KCITTPSPPTTTTTLPPTTTPSPPTTTTTTLPPTTTPTTTTPSPPTTTTTTPLPTTT 1453

Qy   337 PKEPAPTTKSAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKSAP-TTPKEPA 395
               ||: |||| |: ||| |: || : ||||||| ||
Db   1454 PSPPISTTTTTPPPTTTPSP-PTTTTPSPPTTTPSPPT-TTTTPPPTTPIPTTP 1511

Qy   396 PTTPKKPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPAPKPAPTTPK 455
               |||| ||||||||| | | |||||||||||: |: :
Db   1512 ASTTTLPPTTTPSPPTTTTTTPPTTTTTPSPPTTTPITP-PTSTTTLPPTTTPSPPTTTTTT 1570

Qy   456 EPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAP-TTTKEPAPTTTKSAP-TTPKEPSP 513
               |||| |||||::: | |: |: ||||||||||||| ||
Db   1571 TPPPTTTTPSPPTTTTPSPPTITTTTTPPPTTTPPPTTTTTTTTPPPTTTTPIPPT 1630

Qy   514 TTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTKKPAPTAPKEPAPTTPK 573
               :|| ||| | ||| || | ||| ||| ||| |||| |||
Db   1631 STTTLPPTTTPSPPPTTTPPPTTTPSPPTTTTPSPPITTTTPPTTTTPSSPITTTP- 1689

Qy   574 ETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAP 633
               | ||||| | ||:|| ||: |:| :
Db   1690 -SPPTT----TMTTPS--PTTTPSSPITTTTTPSSTTP-SPPPTTMTTPSPTTTPSPPT 1741

Qy   634 NTPKEPAPTTPKEPAPTTP-----KETAPTTP------KG------ 670
               |||||||||||||: || :||||||
Db   1742 TTMTTLPPTTTSSPLTTTPLPSITPTPTTPTTTPTPCVPLCNWTGWLDSGKPNFHK 1801

Qy   671 -- 670

Db   1802 PGGDTELIGDVCGPGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGVI 1861

Qy   671 ----------------TAPTLKEPAPTPKKPAPKELAPTT---TKEPISTTSDKP- 708
                              : |||: : :|||:
Db   1862 PMAFCLNVEINVQCCECVTQPTTM---TTTTENPTPTPITTTTTVTPTPTPTSTQSPNG 1918

Qy   709 -APT-TPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT 765
               | || | ||||||||||||| | || ||||||
Db   1919 LQAPTTPISTTTVTPTPTPTGTQTPT-TTPITTTTTVTPTPTPTGTQTPTTVLITTTT 1977

Qy   766 TKGPTST-TSDKPAPTTPKETAPTTPKEPAPTTPKKP------APTTPETPPPTTSEVS 817
                || |: |||| |: ||||||| ||||: : :
Db   1978 TMTPTPTPTSTKSTTVTPTTTTTTVTATPTPTGTQTPTMIPISTTTTVTPTPTPTTGTSTG 2037

Qy   818 TPTTTK-----EPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTT 872
               : : : ||:: |:|:| :   : |::
Db   2038 PPTHTSTAPIAELTTSNPPPESSTPQTSRSTSSPLTESTTLLSTLPPAIEMTSTAPPSTP 2097

Qy   873 TAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKI 932
               || |: ::||:|| | |||| : : || || || : |
Db   2098 TAPTTTSGGHTLSPPPSTTTSPPGTPTRGTTT--GSSSAPTPSVQTTTTSAWTPTPTPL 2155

Qy   933 TT---LKTTTLAPKVTTTKKTITTTEIMNKPEE--------------- 962
               || : | :| :   : :
Db   2156 STPSIIRTTGLRPYPSSVLICCVLNDTYYAPGEEVYNGTYGDTCYFVNCSLSCTLEFYNW 2215

Qy   963 --TAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPE 1020
               :::: : :   : :| : :
Db   2216 SCPSTPSPPTTPSK-STPTPSKPSSTPSKPPTPGTKPPECPFDPPR----------Q 2261

Qy   1021 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNP------------KSEDAGG 1058
               | : || |: : :||
Db   2262 ENETWWLCDCFM-ATCKYNNTVEIVKVECEPPPMPTCSNGLQPVRVEDPDGCCWHWECDC 2320

Qy   1059 ---AEGETPHMLL------------------------RPHVFMPEVTPDMDYLPRVPNQGIIIN 1095
                  |:||: ||
Db   2321 YCTGWGD-PHYVTFDGLYYSYQGNCTYLVEEISPSVD------NFGVYID 2364

RESULT 2
T18262
s-layer protein - Clostridium thermocellum
C;Species: Clostridium thermocellum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18262
R;Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A;Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulo
e.
A;Reference number: Z18847; MUID:93209931
A;Accession: T18262
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1664 <FUJ>
A;Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match   13.1%;   Score 950; DB 2; Length 1664;
Best Local Similarity 31.0%;   Pred. No. 1.5e-32;

```
cyst germination specific acidic repeat protein precursor - Phytophthora infestans
C;Species: Phytophthora infestans (potato late blight agent)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31108
R;Goernhardt, B.
  submitted to the EMBL Data Library, April 1998
A;Reference number: Z20986
A;Accession: T31108
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1489 <GOE>
A;Cross-references: EMBL:AF061185; NID:g3851513; PIDN:AAC72308.1
C;Genetics:
A;Gene: car90
```

```
Query Match   11.9%;   Score 863.5;   DB 2;   Length 1489;
Best Local Similarity 31.7%;   Pred. No. 5.3e-29;
Matches 368;   Conservative 55;   Mismatches 509;   Indels 229;   Gaps 51;

Qy   104 TKKVIESEEITEEHSVSENQESSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK  163
Db   282 TPTVGYSTEETEGQHVTGGYEPSDETAPTEGTTY---VPREETTAAPSE---DTTYAP    334
Qy   164 KNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTT-----QHNKVSTSPKITTAKPI  217
Db   335 REVTPYAPTEKPYDVEETTYVTEESTYAPTKSETNAPTERMHYAHIEKPCDTEVTMYAPT 394
Qy   218 NPRPSLP--------------PNSDT----SKETSLTVNKETTVETKETT---TNKQT   255
Db   395 EETTYAPTEETTYAPTEETTYAPTEETTYPTEETTYAPTEETTYAPTEKTT         454
Qy   256 STDGKEKTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPK  315
Db   455 YAPTEETTYAPTEETPYEPTEETTYAPTEETTY--EETTYASTEETTYAPTEETTY     511
Qy   316 EPASTPKEPTPTTIKSAPTTPKEPAPT-TTKSAPTTPKEPAPT-TTKEPAPTTPKEPA   372
Db   512 APAEETPYEPTEET-TYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPT        570
Qy   373 PTTTKEPAPTTTKSAPT------------TPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTT 422
Db   571 EETTYAPTEETT-YAPTEETMYAPTEETTYAPAEETPYEPTEETTYAPTEET        629
Qy   423 PKEP------------AP--TTKEPAPTTPKEP-------APTAPKKPAPTTPKEPAP   459
Db   630 TYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAP        689
Qy   460 TTPKEPAPT--TTKEPSPTTPKEP------APT-TTKSAPT-------TTKEP-       496
Db   690 TEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYASTEETTYGPTEET        749
Qy   497 --APT-TTKSAPT--TPKEPSPTTTKEP------APTTPKEPAPTTPKKPAPTTPKEPAP 545
Db   750 TYAPTEATTYAPTEETPYEPTGETTYAPTEETTYAPTEETTYAP                809
Qy   546 T--TPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPKEPAPTT 603
Db   810 TEETPYEPTEETTYAPTEETT-TYTPTEETTYAPTEETTYAPTEKTTYAPTE        867
Qy   604 PEELAPT--TPEEPTPTTPEEP------APT-------TPKAAAPNTPKEPAPTTPK    645
Db   868 ETTYAPTEETPYEPTEETTYAPTKETTYAPTEETTYASTEETTYAPAEETPY        927
Qy   646 EP-----APTTPKEPAPTTPKGTAPT--TLKEPAPTTPKPAPK-ELAPT-         695
Db   928 EPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPTE        987
Qy   696 -TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEP------APTTPKGTAPT----  743
Db   988 ETTYAPTEETMYAPIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTY        1047
Qy   744 ------------TLKEPAPTTPKKP------APKE--LAPT--TTKGPTSTTSDKPAP   779
```

```
Matches 312;   Conservative 104;   Mismatches 361;   Indels 230;   Gaps 55;

Qy   286 VLAKPTP-KAETTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTT  344
Db   758 VVIQAPIKAASDEPIPTDTPSDEPTPS------------DEPTPS--DEPTPSDEPTPSD 804
Qy   345 TKSAPTTPKEPAPTTPKEPAPTTPKEPA----PTTKEPAPTTKSAPTTPKEPAPTTPK  400
Db   805 EPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTP       864
Qy   401 KPAPTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPA-TPKEPAPTAPKKPAPTTPKEPAP  459
Db   865 SDEPTPSDEPTPS--DEPTP-SDEPTPSETPEEPIPTDTPSDEPTPSDEPTP         919
Qy   460 TTPKEPA--PTTTKEPSPT-TPKEPAPTTTKS-APTTTKEPAPT---TTKSAPTTPKEPS 512
Db   920 SDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPT             979
Qy   513 ----PTTTKEPAPT-TPKEPAPTTPKKPAPTTPKEPAPTTPKKPAPTAPKEP        567
Db   980 PSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEP      1039
Qy   568 APT---TPKETAPTTPKKLTPTTPEKLAPTTPEKPAPT--TPEELAPTTPEEPTPT-TP  620
Db  1040 TPSDEPTPSE----TPEEPIPTDTPSDEPTPSDEPTPSD-EPTPSDEPTPSETP      1094
Qy   621 EEPAPTTPKAAAPNTPKEPAPTTPKEPA--PTTPKEPAPT-TPKETAPTTPKGTAPTTLK 677
Db  1095 EEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSD      1154
Qy   678 EPAPTTPKPAPKELAPTTTKEPT-STTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTT  736
Db  1155 EPTPS--DEPTPSD-EPTPSDEPTPSETPEEPIPTDTPSDEPTPS--DEPTPS-      1208
Qy   737 PKGTAPTTLKEPAPT-TPKKPAPKELAPTTTKGPTSTTSDKPAPT---TPKETAPTTPKE 792
Db  1209 ---DEPTPSDEPTPSETPEEPI----------PTDTPSDEPTPSDEPT-EPTPSDE    1253
Qy   793 PAPTTPKKPAPT-TPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENS 851
Db  1254 PTPS--DEPTPSETPEEPIPTDTPSDEPTPSDEPT----PSDEPTP--SDEPTPSDEPTP 1305
Qy   852 PKEPGVPTTKTPAAT--KPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTEKTTE 909
Db  1306 SDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSD----EPTPSDEPTPSDEPTPS      1361
Qy   910 SKITATTTQVTSTTTQDTTPFKITTLKTTLAPKVTTTKTITTTEIMNKPEETAKPKDR   969
Db  1362 DEPTPSDEPTPSETPEEPTP------TTTPTPTPSTT----------PTSG         1396
Qy   970 ATNSKAT------------TPKPQKPTKAP-KKPTSTKPKTMPRVRKPKTTPTPRKM------ 1014
Db  1397 SGGSGGSGGGGGGGGGTVPTSPTTPSSKPTSTPAP---TEIEEPTPSDVPGAIGGEHRA 1453
Qy  1015 ------TSTMPELNPTSRIAEAAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHM- 1066
Db  1454 YLRGYPDGSRPERNITRAEAAVIF------------AKLL-----GADESYGAQSASPYSD 1498
Qy  1067 LLRPH----------------VFMPE-------VTPDMDYLPRVPNQGI-----       1092
Db  1499 LADTHWAAWAIKFATSQGLFKGYPDGTFKFPDQNITRAEFATVVLHFLTKVKGQEIMSKLA 1558
Qy  1093 ---IINPMLSDETNICNG----KPVDGLTTL---RNGTLVAFRGHYFWMLSPFSPPS   1139
Db  1559 TIDISNPKFDD----CVGHWAQEFIEKLTSLGYISGYPDGT-------FKFQN      1600
Qy  1140 PARRITEVWGIPSPIDTVPTRCNCEGKTFFFKD-SQYWRFTNDIKD              1184
Db  1601 YIKRSESV----ALINRALERGPLNGAPKLFPDVNESYWAF-GDIMD             1642
```

```
RESULT   3
T31108
```

```
Db   1048  ASTEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE  1107
                 ||   | |||    || |||  ||     ||  ||||   || ||
Qy    780  TTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTTKEPT--TIHKSPDESTP  837
                ||  | |||  | || ||  | | |:||      |  | |:|:|
Db   1108  ETTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPIEETTYGPTEETTY-APTEATT  1166
                 ||| || ||    || ||| ||
Qy    838  ELSAEPTPKA-LENSPKEPG-----VPTTTKPAATKPEMTTTAKDKTTERDLRTTP----  887
                 ||| |  |  || |   | ||  | | |   :||  | | |
Db   1167  YAPTEETTYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTETPYEPT  1226
                ||||| ||| || |  ||||  || ||||  ||| | |||||
Qy    888  ETTTAAPKMTKETATTTEKTTESKITATTTQVISTTTQDTT---------PFKITT---  934
                ||  | | |   |||   || | || || ||:|:||
Db   1227  EETTYAP-----TEETTYEPTEETTYAPTEETTYAPTEETTYAPIDETTYGP  1281
                |:|  ||      |
Qy    935  LKTTTLAPKVTTTKKTTITTTEIMNKPEETAKPKDRAT--NSKATTPKP-QKPTKAPKKPT  991
                ||||||||||| |  |   |||| |   || |   | | |||
Db   1282  TEETTYAPTEATTYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPMEET  1341
                ||| | |  | |||
Qy    992  -----STKKP-KTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRP  1038
                 ||| |   | |  | | | |    |:|     | | | | |
Db   1342  PYEPAEESTSTVSTEKPCNTEEFTDEPTDEPTDE--PSDEPTDEPTDLPTDEPSTP  1399
                 |: |     | | | |    | || | ||
Qy   1039  --NQTPNSKLVEVNPKSEDAG  1057
                || || | |    :|
Db   1400  CDNQGLNGIGVENKVRYNNAG  1420
```

RESULT   4
T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C;Accession: T16251
R;Leimbach, D.
submitted to the EMBL Data Library, January 1996
A;Description: The sequence of C. elegans cosmid F35A5.
A;Reference number: Z18485
A;Accession: T16251
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <LEI>
A;Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GN00
A;Experimental source: strain Bristol N2; clone F35A5
C;Genetics:
A;Gene: CESP:F35A5.1
A;Map position: X
A;Introns: 1272/2

```
Query Match           11.6%;   Score 846;  DB 2;  Length 1274;
Best Local Similarity 27.6%;   Pred. No. 2.4e-28;
Matches     313;  Conservative 119;  Mismatches  455;  Indels 248;  Gaps  60;

Qy     73  PSSKKAPPPSGASQTIKSTTKRSP----KPP---NKKKTKKVIESEEITEEHSVSENQES  125
               || | || |   |    |      |||    : ||  | |:|     :|
Db    125  SSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNK-KNRTKK-KP------TP-----  173

Qy    185  PSPKKAAPSKEHDPIVPPTPIKNPAKKWKPPWEDDEVPTEEIKEPEPATRKVPALKKKEP  244
               : |||| ||            : |||||||||| |  | |||||  | | :|
Db    245  STSVKPVSDPSPTKKV-----------PVKKEPEVPPTPIKNPTKKWKPPWEDETPVEEVK  294

Qy    126  KPPVVDEAGSGLDNGD---FKVTTPDTSTTQHNKVSTSPKITTAKPI-NPRPSLPPNSD  173
               |||| ::       |||| :   |  | :|||||||||
Db    174  KPPVVDEAGSGLDNGD---FKVTTPDTSTTQHNKVSTSPKITTAKPI-NPRPSLPPNSD  228

Qy    295  EPPVPEKKAPVLKKKDPAPAAKARDPSPSKAAPKKVEPSSPVVPPTPVKNPVKKYKPPWE  354
               || | |||| || || | : |   ||| | |||
Db    229  TSKETSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLA  288

Qy    229  TSKETSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLA  288
               || || | | |  |     |||| |   |||||||
Db    355  VDDEPAEEVKKPSAPEKKTPVLKRKREPEPSSTTPSSSDPSPKKAAPAVKPRDSSPKKATPL  414

Qy    289  KPTPKAE---TTTKGPA  323
               |||  |     | ||
Db    415  QADPKAQEVPPTPVKNPVKYKYKPPWEVDDEDPVEEVKQPEAPAKKTPVLKRKEPAAKDTA  474
```

```
Qy    324  EP----TP-TTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTTKEPA----------  364
                :|      || |||||||| |  | ||:|| || | ||
Db    475  KPATSKTPETPEKKDPVKPRDSPKKVAAKPDSAQAPA-TPVKNPVKKRRPPWEDDETPA  533

Qy    365  -----PTTPKEPAPTTTKEPAPT----------TTKSAPTTPKEP-----APTTPKKPAPTTP  407
               :    |||| ||| |             || |||   ||   || |||
Db    534  DDVSKPTDAKKTPSLAKKDPAPAKESLKPKADTKAPAKPRDPSPKKVAPTAPEKKTPVLA  593

Qy    408  KE---PAPTTPKEPTPTTPKEPAPTTKEPAPTTPK-EPAPTAPKKPAP-------TTPKE  456
               |:    |||     ||   ||  | |||   || | |||| ||||
Db    594  KKEPAGPADSKTKEPEKSKPRDPSPKKAVPAKPVPKTEVAPAVKKPEPISKPDTAPKK  653

Qy    457  PAPTTPKEPAPTTTKEP----SPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTP-KEP  511
               |||||||| |||| ||     ||||||||||||||  ||   |||| |||:  ||
Db    654  AEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPEPEKK-TPVLAKKAPTKPDSEA  711

Qy    512  SPTTTKEPAPTTPK--EPAPTTPKKPAP--TTPKEPAPTTPKEPAPTTTKKPAPTA-PKE  566
               ||| | |||||||  || ||||  |   |||||||||||  || ||| ::
Db    712  AADPVSGPSSKDPKLAKKAPVKPRDPSPMKAVPIKPAPKT-EVPPAVVKKPEPVAKSRD  769

Qy    567  PAPTTPKETAPTTPKKLIPTTP-----EKLAPTTPKEPAPTTPEELAPTTPEEP-----T  616
               |||||| |  | ||  ||||||      ||||| |||| |||   |||||  | ||
Db    770  PSPKKAK-AEPNSP--VVPPTPVKNPVKWKPPWEDDDAPAEPVNVPEPEKKTPVLAKKT  826

Qy    617  PTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTT-PKEPAPTTPKETAPTTPKGTAPTT  675
               ||| |||  | | ||  |||||||||||||||||| |||  || |:||||||  ||||
Db    827  PVKPRDSPSPKKAVPAKPSTKTDAPPVSVKPKEPSPSPKKAEPNSPVVP----PTP-882

Qy    676  LKEPAPTTPKKPAPKELAPT-TTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAP  734
               |||||| ||||  ||   |  :||| || |:
Db    883  VKNPVKKW-KPPWEDDDEPTEEVKKPSE-PEKKTPVLAK-KEPEKPKD-APKVAAKPRD  937

Qy    735  TTPKGTAPTTLKEPAPT-------TPKKPAP----TPKKPAP-----KELAPTTKGPTS----TTSDK  776
               ||  |  |||| |:||        |||| |||       |:|||||||||  ||    :|
Db    938  PSPKKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEAPPTPVKNPVKKWKPPWEDDDE  995

Qy    777  P------APTTPKET-----------APTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTTKEP  825
               |:     ||||||||            |||| |||||||||||||
Db    996  PAEPVSAPEPEKKTPVLAKKAPAKPRDP---SPKKAAPVAAK-PDPKIPEV-PPTPVKNP  1050

Qy    826  TTIHKSP----DESTPELSA-EP---TPKALENSPKEPGV----------PTTKTPA  864
               ||||        || ||  |   | || |||      |:|| ::
Db   1051  VKKWKPPWEDDDEPSEPVSAPEEKKTPVLAKKAPTKPATKPDSEAAADPVSGPTSKDPK  1110

Qy    865  AT------KPEMTTTAKDKTTERDLRTTPETT--TAAPKMTKETATTTEKTTESKITATT  916
               ||       |   | | | || || |  ||    || |: ||||| |
Db   1111  LSKKAPVEKPKPTTDPKDDKLKPSPAKKPEKAPEAAPKKWKPVWDDDPDPEADFTVPA  1170

Qy    917  TQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKAT  976
               :||  | ||
Db   1171  PSKKPDTEDPADPLG---------GPKTKDPK--------LNKKAPAEKPTEK--------  1206

Qy    977  TPKPQKPTKAPKKPTSTKKPKTMPRVRKPK----------TTPTPRKMTSTMP  1019
               |||  |: :|| |||  || || ||
Db   1207  -PKPKEVSKEPPKPTEPPKP-AAPKKWKPPWEDDPDPEADFTMPAPKKPDTEDP  1259
```

RESULT   5
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N;Alternate names: alpha-NAC protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C;Accession: T30826
R;Yotov, W.V., St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A;Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a mus
A;Reference number: Z20889; MUID:96312450
A;Accession: T30826
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2187 <YOT>

A;Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1
C;Genetics:
A;Gene: Naca
A;Map position: 10
A;Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A;Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C;Keywords: alternative splicing; DNA binding; transcription factor

```
Query Match           11.2%;   Score 812;  DB 2;  Length 2187;
Best Local Similarity 27.4%;   Pred. No. 1.1e-26;
Matches 311;  Conservative 128;  Mismatches 449;  Indels 246;  Gaps 51;

Qy    68 NPTSPPSSKKAPPPSGASQTIKSTTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSS 127
            :||||||||||| ||||      ::   |  |  :   :||  ||||
Db   746 SPT-PPSSKGAPVPSTGA----------------PSPSKGAPIVPTESSISSKQVPAEILPSPQ 792

Qy   128 SSSSSSSSSTIWKIKSSKNSAANRELQKLLKVKDNKKNRTKKKPTPKPPVVDEAGSGLDN 187
           :  :||       :||   :|||||        |  VTPTSPKKTSATAVPK------------ 833
Db   793 KTPEVTASRLISAVQSPKVDPIMSD------------

Qy   188 GDFKVTTPDTSTTQHNK------VSTSPKITTAKPINPRPSLPPNSDTSKETSL--TVNKE 240
            || ||||| |  |    ||||   |||||
Db   834 ----DTSATLSLKSVPAVTSLSPPKAPVAPSNEATIVPTEIPTSLKNALAAATPKE 885

Qy   241 T------TVETKETTTTNKQTSTDGKEKTSAKETQSIEKTSAKDLAPTSKVLAKPTPKA 294
          |                     :  ::||: |
Db   886 TLATSTPKVTSPSPQKTPKSVSLKGAPAMTSKKAT---EIAASKDVSPSQ--FPKEVPLL 940

Qy   295 ETTTKGPALTTPKEPTPTTPKEPASTTPKEPTTTIKSAPTTPKE-PAPTTTKSAPTTPK 353
           |    :| |||||||||||||||||  |||||   |||  |||||
Db   941 QHV---PPTSPPKSPVSDTLSGALTSPPPKGPPATLAETPTYPKKSPKPAASKKTPATPS 997

Qy   354 --------EPAPTTTKEPAPTTPKEPAPTTTKEPAPTT--TKSAPT------------ 389
                  :|||||||||||||  | ||   ||:|||
Db   998 PEGVTAVPLEIPPCSKKAPKTAAPKESSATSSSKRAPKTAVSKEIPSKGVTAVPLEISLP 1057

Qy   390 ------------TPKEPAPTTPKK-PAPTTPKEPAPTTPKEPTTPKEPTTPKEPATTKEPAPTTPKE 440
                     ||| :||  |   :||
Db  1058 LKETSKSATPGEKSASSPKRSPKTAGPKE---TPPGGVTAVPPEISLPPKETPQNATPNE 1114

Qy   441 PAPTAPKKPAPTT--PKEPAP-----TTPKEPAPTTTKEPSPTTPK-----EPAPTTTKSA 489
            | |     |||      :||||||||||
Db  1115 SLAASSQKRSPKTSVPKETPPGGVTAMPLEIPSAPQKAPKTAVPKQIPFPEDAVTILAGS 1174

Qy   490 PTTTKEPAPTTT-KSAPTTPK-----------EPSPTTTKEPAPTTPKEPAPTT- 539
           ||||||||||  :|||||:   | :::|||| |  || |||
Db  1175 PLSPKKASKTAAPKEAPATPSVGVIAVSGEISPSPKKTSKTAAPKENSATLPPKRSPKTA 1234

Qy   540 -PKE-------------PAPTTPKEPAP----------TTTKKPAPTAPKEPAPTTPKETAPTT 579
            |||               :|||| |    |   :|||  |||
Db  1235 APKEPATPSSEGVTAVPSEISPSPPTPASKGVPVTLTPKGAPNALAE-SPASPKKVPKTA 1293

Qy   580 PKKLTPTTPEKLAPTTPEK-PAPTTPEELAPTTPEEPTPTT------PEEPAPTTPKA 630
            ||||             :|    :    ||
Db  1294 APEETSTTP------SPQKIPKVAGPKEASATPPSKKTPKTAVPKETSAPSEGVTAVPLE 1347

Qy   631 AAPNTPKEPAPTTPKE-PAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTP----- 684
            |    :| ||| ||||||||||   ||   |||    ||||
Db  1348 IPPSPRKAPKTAAPKETPAPS---PEGATTAPVQIPSPSRKGSKKAGSKE-TPTTPSPEGV 1404

Qy   685 -------------KKPAPKELAPTTTKEPTS------TTSDKPAPTTPKGTAPTTPKE 723
                    :|||    |:|:   :   ||||||||||||
Db  1405 TAAPLEIPISSKKTSKMASPKETLVTPSSKKLSQTVGPKETSLEGATAVPLEIPPSHKKA 1464

Qy   724 PAPTTPKE-PAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTP 782
           ||||||  ||||:|:   |:|:|| :  ||   |:|  |:|   ||||   |||
Db  1465 PKTVDPKQVPLTPSPK-DAPTTLAE-SPSSPKK-APKTAAPPSER-VTTVPPEKPA-TPQ 1519

Qy   783 KETAPTTPKEPAPTTPKKPAPTTPETP-----PPTTSEVSTPTTTK-----EPTTIHKS 831
           |:  |      | :  | |||            |:||| |||||     ||||| ||
Db  1520 KASGTTASKVPVPAETQEVAVSSRETPVTPAVPPVVNPSSHKKTSKTIELKEAPATLPPS 1579
```

Qy   832 PDESTPELSAEPTPKALENSPKE-PGVPTTKTPAAPKEMTTAKDKTTERDLRTTPETT 890
           | :   : || ||| ::| || ||| |::||
Db 1580 PTKSPKIPSSKKAPRT--SAPKEFPASPSIK-PVTT--SLAQTAPPSLQKAPSTTIPKEN 1634

Qy   891 TAAPKMTKETATTTEKTTESKITTATTQVTSTTQDTTPFKITTLKTTTLAPKVTTTKKT 950
           ||| ||| |||| ||:||||||
Db 1635 LAAPAV----LPVSSKSPAAPARASALSPATAAPQTAPKEATTIPSCKKAAATETPIET 1690

Qy   951 ITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRKPKTTPT 1010
           | ||:| |||| ::| | | :||: | || :|:|
Db 1691 STAPSLEGAPKTSE---TSVSKVLMSSP-----PKKASSSKRASTLP----------- 1730

Qy 1011 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETP 1064
            ||| : |||: |:|
Db 1731 ---ATTLPSLKEASVLS---------PTATSSGKDSHISPVS-DACSTGTTTP 1770

RESULT
6
S48478

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR01
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
A;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009;
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA>
A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacc
A;Reference number: S27281; MUID:89031230
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseud
A;Reference number: JC6123; MUID:96323237
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match   11.8%;   Score 809; DB 1; Length 1367;
Best Local Similarity 27.5%;   Pred. No. 9e-27;
Matches 328; Conservative 112; Mismatches 526; Indels 226; Gaps 53;

aa2.rpr

```
Qy  1060 EGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLTLLRNG 1119
              :| ||   |        ::|     |:: |           :|
Db  1110 SPKTVTTTVPCSTGTGEYTTEATTLVTTAVTTVVTTESSTGTNSA-GKTTTGYTTKSVP   1168

Qy  1120 TLVAFRGHYFWMLSPFSPPSPARRITEVWGIPSPIDTVFTRC----NCEGKT 1167
              :   | | :    :||    ::| :|
Db  1169 TT------YVTTLAPSAPVTPATN----AVPTTITT--TECSAATNAAGET   1207
```

RESULT 7

S49915

extensin-like protein - maize

C;Species: Zea mays (maize)

C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999

C;Accession: S49915

R;Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A;Description: Pex genes: pollen-specific genes with extensin-like domains.

A;Reference number: S49915

A;Accession: S49915

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1188 <RUB>

A;Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

```
Query Match           10.8%;   Score 784;   DB 2;   Length 1188;
Best Local Similarity 29.9%;   Pred. No. 8.6e-26;
Matches  230;  Conservative  59;  Mismatches  359;  Indels  122;  Gaps  33;
```

```
Qy   171 PTTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL---PPN 226
Db   460 PTPHSPPAD--------DYVPPTPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQ   510

Qy   227 SDTSKETSLIVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV 286
Db   511 APVG----------------SPPPPVKTTSPPAPIG---SPSPPPVSVV           541

Qy   287 ----LAKPTTKGPALTTPKEPTTTPKEPASTTPKEPTTTPKEPA               341
Db   542 SPPPPVKSPPPPAPVGSPPPPEKSPPPPAPVASPPPPVKSP--PPPTLVASPPPPVKSPP 599

Qy   342 PTTTKSAPTTP--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT-KEPAPTT        398
Db   600 PPAPVASPPPPVKSPPPPTPVASPPPPAPVASPPPMKSPPPPTPVSSPPPPEKSPPPPP  659

Qy   399 PKKPAPTTPKEPAPTT------PKE--PTPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKP 449
Db   660 PAKSTPPPEEYTPPPTSVKSPPPPEKSLPPPTLIPSPPPQEKPTPPSTPSKP-PSSPEKP 718

Qy   450 APTTPKEPAPTTPKEPAPTTPKEPSTTPKEPAPTTTKSAPTTTKSAPTTPK         509
Db   719 SP--PKEVSSPPQTP---KSSPPPAPVSSPPPTPVSSPPPALAPVSSPPPVKSSP----  768

Qy   510 EPSPTTTKEPAPTTPKEPAPTTKKPAPTTKEPAPTTTKKPAPTTKKPAPTTKEPAP     569
Db   769 PPAPLSSPPPAPQVKSPPPPVQVSSP-PPAPKSSPPLAPVSSPPPQVEKTSPPPAPLSSPP 827

Qy   570 TTPKETAP----TTP----KKLTPTTPEKLAPTTPEKPAP-----TTPEELAPTTPEEPT 616
Db   828 LAPKSSPPHVVSSPPPAPVVKSSPPPVVKSSPPLTPKPASPPAHVSSPPEVVKPSTP--PA 885

Qy   617 PTTPEEPAPTTPKAAAPNTPKE-PAPTTPKEPAPTTPKETAPTTPKGTAPTTP        675
Db   886 PTTVISP-PSEPKSSPPPTPVSLPPPIVKSSPPPAMVSSP-PMTPKSSPPPVVVSSPPPT  943

Qy   676 LKE---PA----PTTPKK--PAPKELAPTTKE--PTSTTSDKPAPTTPKGTAPTTP     721
Db   944 VKSSPPPAPVSSPPPATPKSSPPPPA---PKSSPPPPAP                     1000

Qy   722 -KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTKPAPKELAPTTTKGPTSTTSDKPAPT   780
Db  1001 MSSPPPPEVVKSPPPPAPVSSPPPPVKSPPPPPV                          1057
```

```
Qy    44 DCDAQCKKYDK---CCPDYESFCAEVHNPTSPPSSKKAPPSGASQTIKSTTKRSPKPPN 100
Db   174 DLSTGCNNYDNQGHSQTDFPGFYWNIDCDNNCCGTKSSTTTSSTSES--STT-------- 223

Qy   101 KKKTKKVIESEEITEEHSVSENQESSSSSSSSSTIWKIKSSKNSAANRELQKKLKVK   160
Db   224 ---TSSTSESSTTTSSTSESSTTTSSSTSSTT-----APATPTTSCTKEKPTP       274

Qy   161 DNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPR 220
Db   275 PTTTSCTKEKPT-PPHHD----------TTPCT---KKKTTTSKTCTKKTTTPV      314

Qy   221 PSLPNSDTSKETSLIVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDL   280
Db   315 PT--PSSSTTESSS------APVPTPSSSTT--------ESSSAPVTSSTTESSS---  353

Qy   281 APTSKVLAKPTTKGPALTTPKEPTTTPKEPTTTPKEPASTTPKEPTTTIKSAPTTPKEP  340
Db   354 APV------PTPSSSTTESSSAPVTSSTTESSSAPVPTSSSTTESSS             407

Qy   341 APTT-----TKSAP----TTPKEPAPTTTKEPAPTTPKEPAP---TTTKEPAPTTTKSA  387
Db   408 APVTSSTTESSSAPVTSSTTESSSAPVTSS----TTESSSAPVTSSTTESSSAPVPTPSS 463

Qy   388 PTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTP---TTPKEPAPTTKEPAPTTPKEPAPT 444
Db   464 STTESSSAPVT---SSTTESSSAPV---PTPSSSTTESSSAPVT---SSTTESSSAPV   512

Qy   445 APKKPAPTTPKEPAPTTPKEPSPTTPKEPAPTTKSAPSTTTKEPAPTTTKSA         504
Db   513 ------PTPSSSTTESSSAPAPTPS----SSTTESSSAPVTSS---TTESSSAPVTPSS  559

Qy   505 PTTPKEPSPTTTKEPAPTTPKEPAPT-TPKKPAPTTPKEPAPTTPKEPAPTTTKKPAPTA 563
Db   560 STTESSSTPVTSS----TTESSSAPVPTPSSSTTESSSAPVT----PSSSTTESSSAPA  611

Qy   564 PKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEKLAPT--TPEEPTTTPEE 622
Db   612 PTPSSSTTESSSAPVT----SSTTESSSAP-VPTPSSSTTESSSAPVTSSTTESSSA   666

Qy   623 PAP------TTPKAAANTPKEPAPTTPKEPAPTTPKEPAPTTTKEPAPT-TPKGTAPTTL 676
Db   667 PVPTPSSSTTESSSAPVT----SSTTESSSAPVT----SSTTESSSAPVTPSSSTTESS  718

Qy   677 KEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAP---TTPKEPA 733
Db   719 SAPVPTPSSSTTESSSAPVPTPS-SSTTESSSAPVTSSTTESSSAPVTSSTTESSSA   777

Qy   734 PT-TPKGTAPTTLKEPAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKE 792
Db   778 PVPTPSSSTTESSSAPVPTPSSSTTESSSVAPVPTPSSSNITSSAPSTPFSSSTESSSV  837

Qy   793 PAP-----TTPKKPAPTTPETPPPTTSEVSTPT----TTKEPTTIHKSPDESTPELSAE  842
Db   838 PVPTPSSSTTESSSAPVSSTTESSSVAPVPTPSSSSNITSSAPSSI---PFSSTTESFST 894

Qy   843 PTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETAT 902
Db   895 GT--TVTPSSSKYPGSQTETSVSSTETTIVPTKTTTSVTTTPSTTITTTTVCSTGTNSAGE 953

Qy   903 TTEKTTESKITAT--TTQVTSTTTQDTTPFKITTLKTTLAPKVTT---TKKTITTT-EI  956
Db   954 TTSGCGSPKTVTTTVTPTTTTVPTTTTSVTTSSTTITTTTVCSTGTNSAGETTSGCSPKTITTTTFVTKNI 1013

Qy   957 MNKPEETAKPKDRATNSKATTPKPQKP----------------TKAPKKPTSTKKPKTM  999
Db  1014 STSFSETA------SESTTTSPTTVTPTTTSVTTSSTTITTTSPTTPVTVVSSTVVTTEYSTKPGGEITTTFVTKNI 1065

Qy  1000 PRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGA 1059
Db  1066 PTTYLTTIAPTP--SVTTVTNFTPTT-ITTTVCSTGT---------NSAGETTSGC    1109
```

```
Qy  781 -TPKETAPTTPKEPAPTTPKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPEL 839
            | ||  :  || .  |||||
Db 1058 KSPPPPAPISSPPPPVKSPPPPAPVS--SPPPPVKSPPPPAPVSSPPPPIKSPPPPAPVS 1115

Qy  840 SAEPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPET 889
            |-|-|
Db 1116 SPPPAP------VKPPSLP-PPAPVSSPPPVVTPAPPKKEEQSLPPPAES 1158


RESULT     8
T18535
high molecular mass nuclear antigen - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18535
R;Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A;Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A;Reference number: 218955; MUID:9803440
A;Accession: T18535
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1151 <SHI>
A;Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

Query Match.
Best Local Similarity   9.6%;   Score 695;  DB 2;  Length 1151;
Matches  304;   Conservative 108;  Mismatches 466;   Indels 324;  Gaps 59;

Qy   64 AEVHNPTSPPSSKK------APPPSGASQPIKSTTKRSPKPNKKKTKKVIESEITEEH 117
Db   39 AELH-PAAPQPPPKWVPIGGAPPPPGTEPT------PPSKPTDGADAAPKASAELTSPP 90

Qy  118 SVSENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNKKNRTKKKFTPKPPV 177
Db   91 PASPSSPPDGPKAPSGAGEAEAGTPPPSQGPAG----------------TPPP-- 126

Qy  178 VDEAGSGLDNGDFKVTTPDTSTTQHNKVSTS--PKITTAKPINPRPSLPPNSDTSKETSL 235
Db  127 -SQGAAGAPKGDGTAQPSGTKSGADGKPAAQDVPKATTA-ATEARP------ASAASP 176

Qy  236 TVNKETTVETKETTTTNK--QTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPK 293
Db  177 TVPKATAEATAVTAASQSAPAKAATDAAAVTAA----SQSAPKATV-EVKPAAAAVAKEAKA 232

Qy  294 AETTTKGPALTTKEPTP-TTPKEPASTTPKEP----TPTTIKSAPTTPKEPATT--- 344
Db  233 VTAAAAAPKATAEAKPAPVTSPTIPCSSAEAKPLTAASPTASKA--TAEAKPVPATASLM 290

Qy  345 ------TKSAPTTPKEPAPTTK---------------------------EPAPTTPKEPAPT 374
Db  291 ATKVTAEAKPAPSPSVPKATTDTKAVTATAPKAGPDVKPAVAVCAEAKPAPPPPPQQLPK 350

Qy  375 TTKEPAPTTTKSAPTT------PKEPA---------PTTPKKPAPTTPKEPAPTTPKEPTPT 421
Db  351 AAAAAAPTGTELKPATAPPHGSPRANSHTVTVTPPNVPRAAAATVP--TAGAVPKASTGT 408

Qy  422 T---PKEPAPTTKEPAPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEP---APTTTKEPS 474
Db  409 TPAAAPQQPVP---KAAPVTPPSPQQAVPRAATAAA----APVTPQQAVTKAATTTNATP 461

Qy  475 PTTPKEPAPTTTKSAPTTTKEPAP--TTTKSAPTTPKEPSPTTTKEP----APTTPKEP 527
Db  462 PPQPIPKAATTTTATPVTPQQPIPKAGTDAAPPPAVPKAPSDGRAATPGVPNAATDPQKP 521

Qy  528 APTTPKKPAPTTPKEPAP---------TTPKEPAPT-TTKKPAPTAPK------EPAPT 570
Db  522 PPTPQSVPSAVTEPKPQPRAAPPSNEATPAVPSPSPNLKSPLPTIPKPVPLMALTPQPV 581

Qy  571 TPK---ETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELA------PTTP------ 612
                | :  || |  :   :  ||   :::|
```

```
Db  582 TAQMVTQLAATKPSPIVPKASPK-ALMTPPPPPGLPRALAAAKLLGLPSSPVASAMHAK 640

Qy  613 --EEPTTTP-------EEPAPTTP--------KAAAPNTPKE--------PAPTTPKEP 647
           | :   ||        | |:
Db  641 VTPRPLPASPVPMAASPASLGPDAARVALATNAASPGAKPEAAGNGTLMAPMGAANTQM 700

Qy  648 APTTPKEPAPTTPKETAPT----------------------TPKGTPAPT------- 674
           || |  ||
Db  701 APIGAGAAQTAPMGAAHTHVSPMGAGGATQMSPTGAANTHMSPIGAGGATQMSPMGAAN 760

Qy  675 TLKEPAPTTPKKPAPKELAPTTTKEP---------TSTTSDKPAPTTPKG--TAPTTPKEP 724
            :|  |:|
Db  761 TQMSPMGATTQMSPMGAAATTQPSPMGAAATQVTATSAGNTMQVSPMGAATPPGTPSVG 820

Qy  725 APTTPKEPAPTTPKGTAPTTLKEP--APTTPKKPAPKELAPTTTKGP----TSTTSDKP- 777
           || |     | |    | |
Db  821 AATTP-QPSPM---GAATTLMSPMGAATTPQ--PSPMGAVTTQPPPMAATNTTQPPPM 872

Qy  778 APTTPKET---APTTPKEP--APTTPETPP--------PTTSEVSTPTTTK 823
           | ||    |  ::| | |   || |
Db  873 AASTPQSTPMGAATTQSPMGAATTQSPMGASTPQAPPTVAGSPTPPPIPPSPTAQT 932

Qy  824 EPTTIHKSPDESTPEL--SAEPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTER 881
           :| |           |  ||
Db  933 SPQPMSKSPPPDPPKAPSAAAQTSPAAHVANASPGV-TAVSPA---PIGVTEASPSADGA 988

Qy  882 DLRTTPETTTAAPKMT-KETATTTEKTTESKITATTTQVTSTTQDTTPFKITTLKTTTL 940
             : || |         | |            | |      :     |
Db  989 RLSPGTAATDGPKASPAATADVTEAATD--VTAAATAVPA---EAAP---------TK 1033

Qy  941 APKVTTTKTITTTEIMNKPEETAKPKDRATNSKATTPKQKPTKAPKKPTSKKKPKTMP 1000
           | :                                                    : |
Db 1034 AKRSSSSSSSSSSSSSSSSSSSSSSDSDSSSSSSESNPASPA--------------P 1078

Qy 1001 RVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAE 1060
                    :      |  |                       :     || |
Db 1079 AVGDGQQQMTPGAAQSVPP------------VTEAAVQ------------EAAAAAAAGAE 1117

Qy 1061 GE 1062
           ||
Db 1118 RE 1119


RESULT     9
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T25697
R;Fulton, B.
submitted to the EMBL Data Library, August 1996
A;Description: The sequence of C. elegans cosmid F16F9
A;Reference number: 220071
A;Accession: T25697
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1229 <FUL>
A;Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9.2
C;Genetics:
A;Experimental source: strain Bristol N2; clone F16F9
A;Gene: CESP:F16F9.2
A;Map position: X
A;Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match
Best Local Similarity    9.4%;   Score 684.5;  DB 2;  Length 1229;
Matches 337;   Conservative 102;  Mismatches 400;  Indels 407;  Gaps 58;

Qy    2 AWKTLPIYLLLLSVFVIQQVSSQELSCKG-RCFESFERGRECCDAQCKKYDKCCPDYE 60
           ||  :       :   |||| |::| :  ||   |::  :|:|    ||  :|  |
Db    4 AW-VVSFAFLILGNVVQSSLLSKTINSDGSRDFKIVKHIKKNTCTSCSC---KCVPD- 56
```

Mon Apr 29 08:35:16 2002

aa2.rpr

```
Db    977 LELLSK------------INNTQISQPKPTDISKTDALSSLISGLIGSFTKAPMAPTI---- 1022
                 |::::|               |||  ||
Qy    997 KTMPRVRKPKTTPTPRKMTSTMPELN----------------------------PTSRIAEA 1030
                 || :|             :| ||
Db   1023 ------------HTTTDAAFVTATEASLNDGSDKKIIDEAQPTDEIRRA 1059


RESULT   10
A35175
mucin 1 precursor, repetitive splice form A [validated] - human
N;Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episia
ncreatic mucin; polymorphic epithelial mucin (PEM)
N;Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precurso
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
C;Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218;
R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene
A;Reference number: A35175; MUID:90202794
A;Accession: A35175
A;Molecule type: mRNA
A;Residues: 1-952,1033-1344 <LIG1>
A;Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124.
A;Experimental source: splice form A
A;Note: GenBank entries HUMEPISIA1 and HUMEPISIA2 present only the amino-and carboxyl
A;Accession: B35175
A;Molecule type: mRNA
A;Residues: 1-19,29-952,1033-1344 <LIG2>
A;Cross-references: GB:M32739; GB:J05288; NID:g182125; PIDN:AAA35806.1; PID:g182129;
A;Experimental source: splice form B
A;Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl
R;Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burc
J. Biol. Chem. 265, 15286-15293, 1990
A;Title: Molecular cloning and expression of human tumor-associated polymorphic epith
A;Reference number: A35886; MUID:90368715
A;Accession: A35886
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-992,1033-1344 <GEN>
A;Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870
A;Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence
R;Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A;Reference number: A35887; MUID:90368716
A;Accession: A35887
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-109,'S',1111-1339,'A',1341-1344 <LAN>
A;Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599
A;Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated se
R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky,
Eur. J. Biochem. 189, 463-473, 1990
A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen
A;Reference number: S10571; MUID:90276413
A;Accession: S10572
A;Molecule type: mRNA
A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A;Cross-references: EMBL:X52229; NID:g37053
R;Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A;Reference number: S40293
A;Accession: S40293
A;Molecule type: mRNA
A;Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <
A;Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
R;Abe, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A;Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associa
A;Reference number: A36735; MUID:90088473
A;Accession: A36735
```

```
Qy    61 SFCAEVHNPTSPPSSKKAPPPSGASQTIKSTTKRSPKPPNKKTKKVIESEEITEEHSVS 120
                  |::|  |||                    |||||||| |
Db    57 --------APSN------PFDVSTTISS-------------------INNDVDIGPSGD 83

Qy   121 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTPKPPVVDE 180
                  ::| |||                          ::|:||
Db    84 SNPTGSSWFQEIEATVGGQTVKSEHNIDSSVEYEKKV------------------ 120

Qy   181 AGSGLDNGDFKVTTPDTSTTQ---HNKVSTSPKITTA-KPINPRPSLPPNSDT-------- 229
                  |:: | ||      | ||:| ||  ||:|
Db   121 --------TTSnDASTTNAPTTGKDSTTPELIITGIVVINSKSESVTDMSTTRFSTTL 169

Qy   230 SKETSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAK 289
                  :: |||  ||:||||||   ||                 |:|||
Db   170 SPTTELLTSPETLVSTDSSTSST-EQTSPDNTTEIASPMETNTTTEATTTSVEPSVSTLAS 228

Qy   290 PTPKAETTTKGPALTTKEPTTPTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 349
                  |  |          ||::| |||| |||    :|||| |
Db   229 ---EDETTVTAIAEST-------TTVIAEVSTTTEPTTT----------AESTTKKST 267

Qy   350 TTPKEPAPTTPKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTP----KKPAPT 405
                  ||:|||       ||| |||
Db   268 T----KAPATTEEPTPTTTEE--VTTTEAETSTTTSSETSTEK---PTTPLIDNKIAGPA 318

Qy   406 TPKEPAPTTPKEP-TPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTPKE 464
                  || |||  |       :|| ||
Db   319 TGK--PETTHFPVTGTTPN-FDTATETPFVAKSEDKMTLSKTAATETTQQTTEVT-DG 372

Qy   465 PAPTTTKE------PSPTTPKEPAPTTTKSAP------------TTTKEPAPT----TTKSAPTTP 508
                  | |||:|           |::|      |      :||||
Db   373 PEKETTKNVSIEIPITTVPLVETTSTTSTRASKESDGFHTTLKLKVTTADSDSTESATTVK 432

Qy   509 KEPSPTTTKE---PAPT------TPK-----EPAPTTPKKPAP------------ 537
                  ||| |||           |||     ||  |:|
Db   433 PPNEETTTKSHVVPKPTKKGTVKVTPKLELSFDEPTEIT--KAPHPGKLLEKKTYHFVLSD 491

Qy   538 ----TTPKE-----------PAPTTPKEPAP---TTTKKPAPTA--PKEPAPTT 571
                  |||             ||| ||||||
Db   492 NFARYSEAKENDDYNHLDYNHYREAKEPTTTEESSTTEEVTTEEPANTGNPPTTENPTT 551

Qy   572 ---PKETAPTTPKKLPTTPEKLA---PTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAP 625
                  |||:|: |                           |||::|||:|
Db   552 TEQPTSTAESTTTALPTTTEQTVTTEEPTTAEKSTATQ----KPTTTQESVST---EKTS 604

Qy   626 TTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKETAPTTPKGTAPTTLKEPAPTTPK 685
                  |:||   | || |||| |||   :|:|     ||  ||||
Db   605 TTKKA---STTEE-PTTTDEPTTTT---ESSTTGKATTPELSTTSEETTTTELKITTE- 655

Qy   686 KPAPKELAPTTTKEPTSTTSDKPAP------TTPKGTAPTTPKEPAPTTPKEPAPTTPKGTA 741
                  ||||||||||:|:|                 ||| |:||||||||||| |
Db   656 ------GSTTTEEPTTTAIFAEASTGIITTDEETTSTTSTPEITSTKE--IVTESAIT 706

Qy   742 PTTLKEPAPTTP---------------------KKPAPKELAPTTTKGP---------769
                  |||||||||||                      ||:|||||||
Db   707 QTSVSVVESSTPRQLPERWKAIVNKFKHNLEVLKEKKRLLKEKESTSTTGSDSSETTTVV 766

Qy   770 -------TSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPE--TPPPTTSE-VSTP 819
                  ||:|                 |:|||||||
Db   767 AENIDEVTTTEKEKVVQTTPITTEKSTTQEETTTTTTTTTEKTSKTTTSKTTTSESATTE 826

Qy   820 TTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAIKPEMTTTAKDKTT 879
                  |||| ||                        |:|
Db   827 TTTSEPST-------TEST----------TVDTSSATTEESSTAAETTTT 859

Qy   880 ERDLRTTEPTTA-----------------------APKMTKET 900
                  | ||||:
Db   860 SAE---TSETTTSESAAFTTGESPENTALQSSSQKSEENESSAEKPGARRDFVPKKHKTT 916

Qy   901 ATTTEKTTESKITATTTQ-VTST---TTQDTTPFKITTLLKTTTLAPKVT---TTKKTTTT 953
                  |  |||| :|      : ||  |  || :||  | |   ||| ||  | |
Db   917 VKPAETTSAVAASTTTTEPTTTTEKSTTLETTTEATTLNEVTGPAFVTGAPVDETTTINT 976

Qy   954 TEIMNKPEETAKPKDRATNSKATTPKPQ---------------------KPITKAPKKPTSTKKP 996
```

A;Molecule type: mRNA
A;Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A;Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543
R;Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H
J. Biochem. 112, 609-615, 1992
A;Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A;Reference number: JX0235; MUID:93123189
A;Accession: PX0066
A;Molecule type: mRNA
A;Residues: 998-1011,'ES',1014-1017;1018-1032,'T',1034-1037;1038-1057 <MAS>
A;Experimental source: gastric carcinoma cell
R;Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A;Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
A;Reference number: S51026; MUID:95080414
A;Contents: annotation
A;Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region ar
C;Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C;Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C;Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C;Genetics:
A;Gene: GDB:MUC1; PUM
A;Cross-references: GDB:120705; OMIM:158340
A;Map position: 1q21-1q23
A;Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C;Superfamily: polymorphic epithelial mucin
C;Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
F;1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
F;1-62/Region: mucin 1 amino-terminal non-repetitive
F;1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
F;1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
F;1-19,29-212,1033-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
F;138-1017/Region: 20-residue repeats (GSTAPPAHGVTSAPDTRPAP)
F;1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F;1245-1272/Domain: transmembrane #status predicted <TRM>
F;1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match   9.2%;   Score 672;   DB 1;   Length 1344;
Best Local Similarity   27.8%;   Pred. No. 4.4e-21;
Matches 310;   Conservative 78;   Mismatches 472;   Indels 256;   Gaps 56;

```
Qy  235 LTVNKETTVETKETTTNK--QTSTDGKEKTSAKETQSIEKTSAKD-LAPTSKVLAKPT 291
            ||| |              :||| :  |  | |: |||  :| |  ||  || |||
Db   15 LTVLTATTAPKPATVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS 74

Qy  292 P-KAETTTKGP--ALTTKEP---------------------TPTTPKEPASTTP--------KE 324
           ||:|| |  |||||||                       |  | |||||  |        |
Db   75 PGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

Qy  325 PTPTT--------IKSAPTTPKEPATTT------------KSAPTTPKEPAPTTTKEP----A 364
         || ||        |||:|||||||    ||    :||:|   |   |     ||||||    |
Db  135 PAPGSTAPPAHGVTSAPDT--RPAPGSTAPPAHGVTSA--RPAPGSTAPPAHGVTSA 190

Qy  365 PTTPKEPAPTTPKEPAPTTTKSAPTTPKEP---------KKPAPTTPKEP--- 410
              ||||  |  | ||||:||||||||||        || :||   |||
Db  191 PDT--RPAPGSTAPPAHGVT-SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 247

Qy  411 --APTTPKEPTPTTP----------KEPAPTTKEP------APTTPKEPAPTAPK--- 447
           :||||||     ||          ||:|   |||      :||||||||  |||
Db  248 TSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 307

Qy  448 KPAPTTPKEPAPTTPK----EPAPTTKEPSPTTPKEPAPTT---TKSAPTTTKEPAPTT 500
         ||:|||||||      ||||  |  |:||  | |:  |  |||   ||:|||||||||
Db  308 TSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 367

Qy  501 TKSAPTTPKEPSPTTTKEP-------APTTPKEPAPTTPK----KPAPTTPKEPAPTTPK- 549
         |||:||   |:| |  |||       :||||||||||||    ||:||||||||||||||
Db  368 T-SAPDT--RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 424
```

```
Qy  550 ---EPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPE- 605
             :||| | ||||  ||  |  || :| ||    |  |  || |    ||||||  ||
Db  425 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 484

Qy  606 ---ELAPTTPEEPTTTP-------EEPAP--TTPKA--AAPNTPKEPAPTTPK- 645
            : |  |:|||| |        |  |    |:||   |  :|||||||||
Db  485 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 544

Qy  646 ---EPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTP-----------KKPAPKEL 692
            :||| |||||| |      |  |:|  |  ||||||||          ||:|||||
Db  545 HGVTSAPDTRPAPGSTAP----PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGST 600

Qy  693 APTTKEPTSTTSDKAP--TTPKGTAPTTPKEPAPTTPKEPAPTTPKG----TAPTTLK 746
         |||:|||:|||||  |  || || :||  | ||  ||  | |||  |    ||| ||
Db  601 AP-PAHGVTSAPDTRPAPGSTAPPAHGVTS----APDTRPAPGSTAPPAHGVTSAPDTR 655

Qy  747 EPAPTTP----------KKPAPKELAPTTKGPTSTTSDKPAP---TTPK--ETAPTTP 790
         ||| ||           ||:|| |  ||| ||:|||||  |    ||    |||||||
Db  656 APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTR 714

Qy  791 KEPAPTTP----------KKPAPTTPETPPPTTSEVSTPTTTKEPTT-----IHKSPD 833
         | ||  |           ||:|| ||:||| || | ||  || ||     |:|||
Db  715 PAPGSTAPPAHGVTSAPDTRPAPGS--TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD 772

Qy  834 ESTPELSAEPTPKALENSPKEPGVPTTKTPAA---------TKPEMTTTAKDK--TTERD 882
         | ||  |  || |   |  | |  | || |             ||||| |||:|   | |
Db  773 TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD 832

Qy  883 LRTTPETT-----------TAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTP-FKIT 933
         |:|| ||            |  || || || |:| |||  || ||  | ||||| | |||
Db  833 TRPAPGSTAPPAHGVTSAP---DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT 888

Qy  934 TLKTTLAPKVTTTKKTITTT--EIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPT 991
         ||:|| || |||  |   |   ||:|  |  |  | |       ||  ||:|   |||
Db  889 SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT-----GSTAPPAHG 946

Qy  992 STKKPKTMPRVRKPKTTPTTPRKMTSTMPELNP---TSRIAEAMLQT-TTRP---NQTPN 1043
              |   |  |    |  |||  ||:| |||     ||  | | |  ||||    ||||
Db  947 VTSAPDTRP---APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 1003

Qy 1044 SKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDETN 1103
          ||:|  | | |    |  ||  |  | |  |||  | |   |::|    |||| |||
Db 1004 AHGVTSAPDTRPAPGSTAPPAH----------GVTSAPDNRPALGSTA---PPVHNVTS 1049

Qy 1104 ICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPS 1139
         || |  |  |   ||| |  |  ::| |||| |||
Db 1050 ASGSASGSASTLVHNGTSARATTTPASKSTPFSIPS 1085
```

RESULT   11
I38346
elastic titin - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: I38346
R;Labeit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A;Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A;Reference number: A57430; MUID:96026330
A;Accession: I38346
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-7962 <RES>
A;Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:127867; OMIM:188840
A;Map position: 2q31-2q31

Query Match   9.2%;   Score 666.5;   DB 2;   Length 7962;
Best Local Similarity   24.2%;   Pred. No. 4.4e-20;
Matches 268;   Conservative 120;   Mismatches 491;   Indels 229;   Gaps 50;

```
Qy     57 PDYESFCAEVHNPTSPPSSKRAPPSGASQTIKSTTKRS--------------PKPNKK 102
             |: :| ||  |::  |   ||               |
Db   6827 PEEEEFVPE--EEVLPEVKPKVPVPAPVPEIKKVVTEKKVVIPKKEEAPPAKVPEVPKKV 6884

Qy    103 KTKKVI-ESEEITEEHSVSENQESSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVK 160
Db   6885 EEKRIILPKEEEVLPVEVTEEPEEPISEE--------------EIPEEPSIEEVEVAPPRVP 6935

Qy    161 DNKNRTKKKPTPKPPVVDEAGSGLDN--GDFKVTTP-DTSTTQHNKVSTSPKITTAKP- 216
Db   6936 EVIKKAVPEAPTPVPKKVEAPPAKVSKKIPEEKVPVPVQKKEAPPAKVPEVPKKVPEKKV 6995

Qy    217 -INPRPSLPP--NSDTSKETSLTVNKETTVETK---ETTTTNKQTSTDGKE--------- 261
Db   6996 LVPKKEAVPPAKGRTVLEEKVSVAFRQEVVVKERLELEVVEAEVEEIPEEEEFHEVEEYF 7055

Qy    262 --------------KTTSAKETQSIEKT------SAKDLAPTSKVLAKPTPKAETTTKG 300
Db   7056 EEGEFEHEVEEFIKLEQHRVEEEHRVEKVHRVIEVFEAEEVEVFEKPKAPP--------KG 7107

Qy    301 PALT----TPKEPTPTTP-KEPASTTPKEPTPTTIKSAPTTPKEP-APTTTKSAPTTPK 353
Db   7108 PEISEKIIPPKKPPTKVVPRKEPPAKVPEVPKKIVVEEKRVPEEPRVPPTKVPEVLPPK 7167

Qy    354 EPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEP----PAPTTPKE 409
Db   7168 EVVP--------------------EKKVPVPPAKKPEAP------PKVPEAPKEVVPPPKKPEV 7213

Qy    410 PAPTTPKEPTPTTPKEPAPTTP------------KEPAPTAPKKPAPTTPKEPAPTTP 462
Db   7214 PPTKVPEVPKVPKAAVPEKKVPEAIPPKPESPPPEVFEEPEESPSADPK-KPEVPPVRVPEVP 7272

Qy    463 KEPAPTTTKEPSPTTPKEPAPTTTKSAPTTKEPAPTTTKEPSPTTTKEPAPT 522
Db   7273 KEVVPEKKVPAAP--PKKPEVTPVK-VPEAPKEVVPEKKVPVP-PPKKPEVPPTK------- 7323

Qy    523 TPKEPAPTTPKKPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKK 582
Db   7324 VPEVPKKVAVPEKKVP---EAIPPKPESPPPEVFEEPEEVALEEPPAEVVEEPEEAAPPQ 7379

Qy    583 LT----PTTPEKLAPTT----PEKPAPTTPEELAPTTPEEPTP-TTPEEPAPTTPKAAAP 633
Db   7380 VTVPPKNPVPEKKAPAVVAKKPELPPVKVPEVPKVPEVPEKKVPLVVPKK--PEAAPPAKVP 7437

Qy    634 NTPKEPAP------TTPKEP------APTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTT 683
Db   7438 EVPKEVVPEKKVAVPPKKPEVPVPPAKVPEVPKKPV---LEEKPAVP---VPERAESPPPEV 7490

Qy    684 PKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKGTAPT 743
Db   7491 YEE--PEEIAP---EEEIAPEEEKPVPVAEE---EEPEVPPPAVPEEPKKIIPEKKVP- 7540

Qy    744 TLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETA------PTTPKEPAPT 796
Db   7541 VIKKPEAPPKEPEPEKV--IEKPKLKPRPPPPPPAPPKEDVKEKIFQLKAIPKKKVPE 7597

Qy    797 TPKKP-APTTPETPPPTTSEVS--TPTTTKEP------TTIHKSPDESTPELSAEPTP 845
Db   7598 NPQVPEKVELTPLKVPGGEKKVRKLLPERKPEPKEEVVLKSVLRKRPEEEEPKVE----P 7653

Qy    846 KALENSPKEPGVPTTKTPAATKP---EMTTTAKDKTTERDLRTTPETTTAAPKMTKETA 901
Db   7654 KKLE-KVKKPAVP---EPPPKPVEEVVPTVTVKRERKIPEPTKVPEIKPAIPLPAPEPK 7709

Qy    902 TTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTEIMNK-- 959
Db   7710 PKPEA------------------EVKTIKPPPVEPEPTPIAAPVTVPVVGKKAE 7745

Qy    960 ------PEETAKPKDRATNSKATTPKPQKFTKAPKKPTS--TKKPKTMPRVRKPKTTPTP- 1011
Db   7746 AKAPKEEAAKPKGPIKGVPKKTPSPIEAERRKLRPGSGGEKPPDEAPFTYQLKAVPLKFV 7805

Qy   1012 --RKMTSTMPELNPTSRIAEAMLQTTT 1036
             :  :|    |:  |  :|: :  ||
Db   7806 KEIKDIILTESEFVGSSAIFECLVSPST 7833
```

RESULT 12

T45025

mucin MUC5B, tracheobronchial [imported] - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 21-Jan-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C;Accession: T45025

R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.

J. Biol. Chem. 272, 3168-3178, 1997

A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter

A;Reference number: Z22899; MUID:97166151

A;Accession: T45025

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-3570 <DES>

A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503

A;Experimental source: placenta

C;Genetics:

A;Gene: MUC5B

Query Match 9.0%; Score 653; DB 2; Length 3570;
Best Local Similarity 22.1%; Pred. No. 7.3e-20;
Matches 394; Conservative 120; Mismatches 588; Indels 680; Gaps 71;

```
Qy     22 VSSQELSCKGRCFESFERGRECDCDAQCKKYDKCCPDYE--SFCAEV-HNPTSPPSSKKA 78
Db   1572 VPLRELGQVVEC--SLDFGLVCRNREQVGKF-KMCFNYELRVFCCNYGHCPSTPATSSTA 1628

Qy     79 PPPS----------------------------GASQTIKSTTKRSRPKPNKKKTKKVIESEEITEEHS 118
Db   1629 TPSSTPGTTWILTEQTTAATTTATTGSTAIPSSTPGTAPP----KVLTSQATTPTAT 1682

Qy    119 VSENQESSS------------------------SSSSSSSSTIWKIKSS----KNSAANR--- 151
Db   1683 SSKATSSSSPRTATTLPVLTSTATKSTATSFTPIPSSLTGTTGTSQNRPHPMAMSTIH 1742

Qy    152 ------------ELQKKLKVDNKKNRTKKKPTPKPP-----VVDE------AGSGLDNG--- 188
Db   1743 PSSTPETTHTSTVLTTKATTTRATSMSTPSSTPGTTWILTELTTAATTTAALPHGTPSS 1802

Qy    189 ------DFKVTTPDTSTTQHN----------------------KVSTSPKITTAKPINPR------ 220
Db   1803 TPGTTWILTEPSTTATVTVPTGSTATASSTRATAGTLKVLTS-TATTPTVISSRATPSSS 1861

Qy    221 -------PSLPPNSDTSKETSLTVNKETTVET----------KETTTTNKQTSTDGKEKTTSAKE 268
Db   1862 PGTATALPALRSTATTPTATSVTATPSSSLGTAWTRLSQTTTPTATMSTATPSSTPETVH 1921

Qy    269 TQSIEKTSAKDLDAPTSKVLAKPTPKAETTTK-------------------GPALTTP----K 307
Db   1922 TSTVLTTTATTTRTGSVATPSSTPGTAHTTKVPTTTTTGFTATPSSSPGTALTPPVWIST 1981

Qy    308 EPTPT-------TPKE--------------------------PASTTPKEPTPTTIKSAPTT 336
Db   1982 TTTPTTRGSTVTPSSIPGTTHTATVLTTTTTTVATGSMATPSSTQTSGTPPSLTTTATT 2041

Qy    337 PKEPAPTTTKSAP--TTPKEPAPTTT-KEPAPT----TPKEPAPTTTKEPAPTTTKSA-- 387
Db   2042 ITATGSTTNPSSTPGTTPIPPVLTTATTPAATSSTVTPSSALGTTHTPPVPNTTATTHG 2101

Qy    388 --PTTP------------------KEPAPTTPKKPAPTT----PKEPAPTTPKEPT 419
Db   2102 RSLPPSSPHTVPTAWTSATSGILGTTHITEPSTPGTSHTPAATTGTTQPSTPALSSPHPSS 2161

Qy    420 PTTPKEPAPTTKEPAPT--TPKEPAPTAPKKPAPTTPKEPAPTT-PKEPAPTTTKEPS-- 474
Db   2162 RTTESPPSPGTTTPGHTRGTSRTTATATPSKTRTSTLLPSSPTSAPITTVVTTGCEPQCA 2221
```

```
Qy    475 ----------------------------------------         474
Db   2222 WSEWLDYSYPMPGPSGGDFDTYSNIRAAGAVCEQPLGLECRAQAQPGVPLRELGQVVEC  2281

Qy    475 ----------------------------------------         475
Db   2282 SLDFGLVCRNREQVGKFKMCFNYEIRVFCCNYGHCPSTPATSSTATTPSSTPGTTWILKL 2341

Qy    498 PTTTKSAPTTPKEPSPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTTKSAPTTT--KBPA 557
Db   2342 TTTATTESTGSTATPSSTQGPPAGTPH--VSTTATTPTVTSSK----ATPSSPGTATA    2395

Qy    558 KPA----------PTAPKEPA-----PTTPKETAPTTPKKLTP---------         585
Db   2396 LPALRSTATTPATSFTAIPSSSLGTTWTRLSQTTTPMATMSTATPSSTPETVHTSTVLT   2455

Qy    586 ----------TTPEKLAPTTPEKPAPTT-----PEELAPTTPEEP-----TPTTPE      621
Db   2456 TTATTGATGSVATPSSTPGTAHTTKVPTTTTTGFTVTPSSSPGTARTPPWISTTTTPT   2515

Qy    622 EPAPTTPKAAANPKPEPAP--TTPKEPAPT----TPKEPAPT--TPKE--TAPTTPKGTA 672
Db   2516 TSGGTVIPSSIPGTTHTPTVLTTTTQPVATGSMATPSSSTQPPSLITTATTITATG     2575

Qy    673 PTTLKEPAPTTPKKPAPKELAPTT-TIKEPTS---------         716
Db   2576 STT-NPSSTPGTTPIPPELTTTATTPAAATSSTVIFPSSALGTTHTPPVPNTTATTHGRSL 2633

Qy    717 APTTP------------------         716
Db   2634 SPSSPHTVRTAMTSARSGTLGTTHITEEPSTGTSHTPAATTGTTTTSTPALSSPHPSSRTT 2693

Qy    753 PKKPAPKELAPTTKGPTSTTSDKPAPTTPKETAPTT----PKEP---------        793
Db   2694 ESPPSP-----GTTTPGHTTATSRTTATNATPSKTRTSTLLPSQPTSAPITTVVTTGCEPQC 2749

Qy    794 --TPAATK--------------         793
Db   2750 AWSEWLDYSYPMPGPSGGDGDTYSNIRAAGAVCEQPLGLECRATAQPGVPLGELGQVVE   2809

Qy    794 --------KEPAPTTPKEPAPTTPKGTAPTTLKEPAP---SE         815
Db   2810 CSLDFGLVCRNREQVGKFKMCFNYEIRVFCCNYGHCPSTPATSSTAMPSSTPGTTWILLFE 2869

Qy    816 VSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTK------         861
Db   2870 LITTATTTASTGSTATP-SSTP--GTAPPKVLISPATTPTATSSKATSSSSPRTATTLP   2926

Qy    862 --TPAATK----PEMTTTPAKDKTTERDLRTTPETTI-TAAPKMKTK         898
Db   2927 VLTSTATLKSTATSVTPIPSSTLGTTGTLPEQTTPVAIMSTIHPSSTPETHTSTVLLTTK  2986

Qy    899 ETATTTEKT----TESKITATTTQVTS-TTTQDTTP----FKITTLKTT         938
Db   2987 ATTRATSSTSTPSSTPGTTWILTELTTAATTTAGTGPTATPSSTPGTTWILELTTTATT   3046

Qy    939 TLAPKVTTT-KKTTTTTEIMNKPEE-----TAKPKDRAT--NSKATTPKPQ------KPTK 985
Db   3047 TASTGSTATLSSTPGTTWILTTEPSTTATAVTAPPGSTATIASSTQATACTPHVSTTATTPTV 3106

Qy    986 APKKPTSTKKPKT----MPRVRKPKTTPTPRKMT---------         1015
Db   3107 TSSKATPSSSPGTATALPALRSTATTPTATSFTAIPSSSLGTTWTRLSQTTTPTATMSTA  3166

Qy   1016 --STMPELNPTSRIAEAMLQTT-----TRPNQTPNSKLVEVNPKSEDAG--GAEGETPH   1065
Db   3167 TPSSTPETVHTSTVLTTTATTTGATGSVATPSSTPGTAHTTKVPTTTTTGFTATPSSSPG  3226

Qy   1066 MLLRPHVFMPEVTPDMDYLPRVPNQGIINPMLSDETNICNGKPVDGLTTLRNGTLVAFR   1125
Db   3227 TALTPPWWISTTTTTPTTTTTP--TTSGSTVTPSSIPGTTHTARVLTTTTTTTVATGSM--- 3280

Qy   1126 GHYFWMLSPFSPPSPARRITEVWGIPSPIDYVFTRCNCEGKT         1167
```

```
Db   3281 ------------------AIPSSS---TQTSGTPPSLITTATTIPATGST   3309
```

RESULT 13
I51618
nucleolar phosphoprotein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C;Accession: I51618; S57757
R;Cairns, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A;Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp
A;Reference number: I51618; MUID:96019267
A;Accession: I51618
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-990 <CAI>
A;Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921
A;Gene: xNopp180
C;Genetics:
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

```
Query Match             8.8%;  Score 643.5;  DB 2;  Length 990;
Best Local Similarity  24.2%;  Pred. No. 5e-20;
Matches  246;  Conservative  148;  Mismatches  424;  Indels  199;  Gaps  40;

Qy     71 SPPSSKKAPPSG-------------------ASQTIKSTTKRSPKPPNKKKT---       116
Db     57 SPDAKRKPPANGLPFKKKSAKESSSSEEDEPPAKKRAQPAGGKKPVVKAVQPKKA       116

Qy    117 HSVSENQESSSSSSSSSSSSSSTIWK-IKSSKNSAANRELQKKLKVKDNKKNRT------KK   169
Db    117 KSSSEDSSDSDSDSEEEETKPPAKRPAQTPKVAAVKTPTQKKAKSSSSEESSSSEDEASKKK  176

Qy    170 KPTPKPP----VVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPPN  226
Db    177 QPVIKVPKQAVVKRAGLASNNG----KIADSSSSEDSDSPPAKKTAATK------TPP-    224

Qy    227 SDPISKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTS--  284
Db    225 --TKPATAAKPQAKKTAGKKSSSREDSSDSDEEQKTAKSKPKPDVSAVP---PPTSVS     279

Qy    285 --KVLAKPTPKAETTTKGPALTTPKEPTPTTKE--PASTTPKEPPTTIKSAPTTPKEP     340
Db    280 KKKTLSQPGTKAKPESSDSDSSDSEEEQPAKAKAKIVPAKAAANSAPRLAKKAETSDSES   339

Qy    341 APTT----------TKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTTKE-PAPTTPKEPAP   388
Db    340 DSSEDEKKSSVKLGVKRAAP---KKAPA-----APDAKSTPVAAAKKSAPAKKASSSSD    390

Qy    389 TTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPA-PTTPKEPAPTTKKPAPTAPK   447
Db    391 SDSSSNEETTTKPAAKTTPAKSA-ATPTSKIPTNGK-ATPTSKTPAKPGTPKTSTAKKDS   448

Qy    448 KPAPTTPKEPAPTTPKEPSPTTPKEPAPTTTKEPAPTTPKEPAPTTKEPAPTTTKSAPTT   507
Db    449 SSSDSDSDSSSDEETTTKPAAKTTPAKSAAATPSKITPTNSKATPTSKKTPAKPGTPKTSAA 508

Qy    508 PKEPSPTTTKEPAPTTPKKPAPTTPKEP-APTTPKEPAPTTPKKPAPTAPKE           566
Db    509 KKDSSSDSSSSDSDSSDEKKTPA-----KRAAKTTPAKPAAAKTTPAKPA------AKTTPAK 557

Qy    567 PAPTTPKETAPTTPKKLTPTTPEKLAPTTPEELAPTTPKEPAPTTPKEPP            626
Db    558 PA------------AKSTPGKVQVPTKKE----SSSSDSDSSDSSSEEKKS           590

Qy    627 TPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKK   686
Db    591 SAKPAVKTT--PGKRATSK---PVVASKPVP-ARKASSSSDSDSDSDSSEEETKITKPLIKLS 643
```

```
Qy    687 PAPKELAPTTTKEPTSTTTSD--------KPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKG 739
          ::                          :::||
Db    644 PAVKTLPPKKAESSSDSSSDSSSDDSEKKTKPAKPPAKSATPVNTKAPAQNKASKASCSDSDS 703

Qy    740 TAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTTSDKPAPTTPKETAPTTPKEPAPTTPK 799
          ::                    :::||
Db    704 SSEEEGKSKQP-TGKSPAAKATAP-PKKNPVAVNKDKPSSSSSSSSSDSSGDDEKQ----KPK 757

Qy    800 KPAPTTPEPTPPPPTTPSEVSTPTTTKEPTTIHKSPDESTPELSAEPTTKRALENSPKEPGVPTT 859
          :|||||||                  :| ||:::
Db    758 QAA------------------AAKDVKQGAKAAAKPTPKKAASSSSE---DS 787

Qy    860 TKTPAATKPEMTTTTAKDKTTERDLRTTPETTTNAAPKMTKETATTTEKTTESKITATTTQV 919
          :||||            |||:|
Db    788 SSDEDVSKAKKTNTAVSKSPV----TTPKAVPAARKKESSSSSDSEDEKQGGKNTSTTKI 843

Qy    920 TSTTTQDTTPFKIITTLKTTTLAPKVTTTK--KITTTTEIMNKPEETA-KPKDRATNS--- 973
          :||||                   ||:
Db    844 ANST--------------PRAAAAECSEESSSSSEDEGKANGTSGKRKRESTGNAEC 885

Qy    974 KATTPKPQKPTKAPKKPTSTKKKPKTMPRVRKPKTTPTP-RKMTSTMPELNPTSRIAE 1029
          :|||                |||:|      :|:|
Db    886 EAVTPE-----NKKLRAKSPNTFPKVNKKELKNTPPRRVVEEDIEINP--RMAD 932
```

RESULT 14
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34513
R;Favello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of C. elegans cosmid ZK783.
A;Reference number: Z21536
A;Accession: T34513
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-3507 <FAV>
A;Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A;Experimental source: strain Bristol N2; clone ZK783
C;Genetics:
A;Gene: CESP:ZK783.1
A;Map position: 3
A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match   8.7%;   Score 636;   DB 2;   Length 3507;
Best Local Similarity 22.6%;   Pred. No. 3.6e-19;
Matches   271;   Conservative   173;   Mismatches   459;   Indels   296;   Gaps   47;

```
Qy     25 QELSCGKGRCFESFERGRECDC------DAQ----------CKKYDKC-----CPDYE 60
          ||:::                     :||           ::
Db   1853 QSLQISGLCAQKNDCNKHAECIDIHPDSHFCSCPDGFIGDGMICDDVDECNNAGMCDDEN 1912

Qy     61 SFCAEVHNPTSPPPSSKKAPPSGASQTIKSTTKRSPKPPNKKKTKKVIESEITEEHSVS 120
          :|||                                     :::|
Db   1913 TKC--------------ENTIGSFNCVCLEGFKKVDEKCVVDEKKQPNREKIE 1951

Qy    121 ENQESSSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTPKPPVVDE 180
          :| ||            :::||:||                              :|||
Db   1952 IDEENSSSSNSG----------QEKPTIKGIVSST 1976

Qy    181 AGSGLDNGDFKVTIPDTSTTQHNKVVSTSPKITTAKPINPRPSLPPN---SDTSKETSLTV 237
          :||                      :|||         |||:
Db   1977 S----ATSSESTTAEPHVTTSISSTTSTTKDMTSKSKSPENVTMSSESPEVSTSS 2025

Qy    238 NKETTV--ETKETTTTNKQTSTDGKEKTTSAKEIQSIEKTISAKDLAP----TSKVLAKPTP 292
          :||     |||                                        :|||
Db   2026 SKSTTASEITVSSTPSESSSSEAPLTSSPATTTEVITESSVKSTTPKEESSEITVKLSS 2085

Qy    293 KAETTTKGPALTTPKEPTPTTPKEPAPTTTIKS-APTTPKEPAPTTTIKSAPTT 351
          ||||:|||                              |||:
```

RESULT 15
T11622
extensin class 1 precursor - cowpea
extensin class 1 precursor - (cowpea)
C;Species: Vigna unguiculata (cowpea)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C;Accession: T11622; S54155
R;Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
Mol. Plant Microbe Interact. 10, 95-101, 1997
A;Title: Rhizobia modulate root-hair-specific expression of extensin genes.
A;Reference number: Z17301; MUID:97155574
A;Accession: T11622
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ

```
Db   2086 KSPEVTESSVKSKSPSTPS--TTSQSVTSTVPETSKSTVLSSEAPVTSTSPTEVHI-SSETK 2143

Qy    352 PKEPAPTTTKEPAPTTPKKPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPK -EP 410
          ::::                  ||||
Db   2144 PSLSASSTTGDTNSTTPSTSSLASVKSTSAPEGTSASVAPVKLSSLSPDVSQPSTKTFDA 2203

Qy    411 APTTPKEPTTPTTPKEPAPTTPKEP-APTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTT 469
          ::|||||                :||||
Db   2204 TESSTVQASEISSGTSVKSTSEPESHVTKLSITSSNPSSSVPVTSPKSTPTVPE----S 2258

Qy    470 TKEPSPTTP--KEPAPTTTKSAPTTPKEPAPTTTKSAP-----TTPKEPSPTTTKKPAP 521
          |||||            :::|
Db   2259 TEQPTSTTPSGQSLTPMNSNSEVLTTSEPHVLSSSLSPDVSQSSTTPNNLSESSTVE-- 2315

Qy    522 TTPKEPAPTTPKPKPAPTTPKEPAPTTPKEPAPTT----------TKKPA 560
          :||||                          ::|
Db   2316 -TPKTSSEVSLNSEEPSTTEAPTTLSPDILSTTTNNLSQSSTVSTEDRSEISSSENSEKPT 2374

Qy    561 PTAPK----------PT------- 610
          :|||
Db   2375 -SAPELVISSVTHVASSSPDVPTES--SEPDDLTGSSTENIPEASSKQTISSTTPPDITTT 2431

Qy    611 TPEEPTPTTPEEP--------------APTTPKAAAPNTPKEPAPTT-------PKEP 647
          :||                                               ::
Db   2432 ASEEPTKSTMSMSPDLSTTSNVLSESSTTPSESSS----KSPVSSSTEGISVVTSTEFSKVP 2487

Qy    648 APTT---------TPKEPAP----TTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELA 693
          ||||
Db   2488 ESTISSVLEEDLIKTTTPSPLIEETTTTASETSEEPLTEDSLTVSVRIHELTTTSSENVPKESE 2547

Qy    694 PTTTPKEPTSTTTSDKPA---------PT--------TP--KGTAP--TTPK- 722
          :|||
Db   2548 STTTSSESSKPSGPEAGILTSTVVVPTSSSVSLITASEIEAITISNTPFKQGRPITTSPKS 2607

Qy    723 ------EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPT-STTSD 775
          :|||
Db   2608 LVKSTTSPSTVTSSEPSESTKRTTVSTTVSTTPTEETTTSESLILIRAAPSKPTESTTES 2667

Qy    776 KPAPTTPKETAPTTPKEPAPTTPKKPAPTTPKEPTPPPTTSEVSTPTTTKEPTTIHKSPDES 835
          :|||
Db   2668 SEAPTTPAKISETKPSNVSSTSRKSTENVETSTSQSGSLESSTMSSTS----SEPETN 2721

Qy    836 TP--ELSAEPTPKALE--NSPKEPGVPTTKTPAATK-----PEMTTTAKDKTTERDLRTT 886
          ::|
Db   2722 APAVTVSSEASSTTLEENSSTSSP----TSSEASVKLSSLFPESTTSEAVTVSSR----A 2773

Qy    887 PETTTNAAPKMTKETATTTEKTTESKITATTT----------QVTSTTTIQDTTP-F 930
          :||
Db   2774 PAEITMSSESHREISTVSSEPSEPEIPLSTTVSPNVVTASSIPSEEPILSSSVTSSSTPRV 2833

Qy    931 KITT-----LKTTTLAPKVTTTKITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTK 985
          :|||
Db   2834 RLITGTPDDLLIVSVTVPSHGNRRQNITASSV---PSNSTSPIILPSES-LTTPQPPPTTT 2889

Qy    986 APKKPTSTKKPKTMPRVRKPK---TTPTPRKMTSTMPELN------PTSRIAEAMLQTTT 1036
          :|||
Db   2890 TTAKPAITSGKRGPPSTQPPAEMFTTPAP------PPPSNGGYGEETNQEEEQVTSTTTT 2942
```

A;Residues: 1-489 <ARS>
A;Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937
A;Experimental source: sub_species Red caloona
R;Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
submitted to the EMBL Data Library, April 1995
A;Description: A class of root-hair specific extensins involved in rhizobium/legume inte
A;Reference number: S54155
A;Accession: S54155
A;Molecule type: mRNA
A;Status: Preliminary
A;Residues: 326-489 <AR2>
A;Cross-references: EMBL:X86030; NID:g791149; PID:g791150
A;Genetics:
A;Gene: Ext26G
C;Superfamily: hydroxyproline-rich glycoprotein
C;Keywords: glycoprotein; hydroxyproline
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-489/Product: extensin class 1 #status predicted <MAT>

Query Match   8.7%; Score 633; DB 2; Length 489;
Best Local Similarity   31.2%; Pred. No. 6.8e-20;
Matches 149; Conservative 33; Mismatches 252; Indels 44; Gaps 7;

```
Qy    337 PKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTPKEPAP 396
          |:|    |  | |      |   |       |||
Db     39 PKQTPPYYVNAPPYYYKSPPP----PSPSPP------------PPPYVHKYPPYYYKSPPP 83

Qy    397 TTPKKPAPTTPKEPAPTTPKEPIPTTPKEPAPTTKEPAPTTP---KEPAPTAPKKPAPTT 453
          | | |   | |  |  |  |||        |  |  |  |     | || | | |
Db     84 PSPSPPPPYYVYKSPPPSPSPPPPYYYKSPPPS--PSPPPYYYKSPPPSPSPPPPYY 141

Qy    454 PKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTTKEPAPTTTKSAPTTPKEPSP 513
          |  |  |  |    |     |  |  |     |   |:|        ||
Db    142 YKSPPPSPSPPPPYYVYKSPPPSPSPPPPYYYKSPPP------------PPSPSPPPP 187

Qy    514 TTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTTKEPAPTTTKKPAPTAPKEPAPTTPK 573
          | | |  |  ||  |   |  |||         |       |:||     |
Db    188 YYYKSPPPSPSPPPPYYVYKSPPPSPSPPPPYYYKSPPPSPSPPPPYYYKSPPPSPS 247

Qy    574 ETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTTPEEPAPTTPKAAAP 633
          |  |  |  | |  |  |  ||      |  ||    |  |:      ||   |
Db    248 PPPPYYVYKSPPPSPSPPPPYYYKSPPPSPSPPPPYYYKSPPPSPSPPPPYYYKSPPP 307

Qy    634 NTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPAPKELA 693
          | | | |  | |||        |  |  |      | | |  |:|
Db    308 PSPSPPPPYYVYKSPPPSPSPPPPYYYKSPPPSPSPPPPYYYKSPPPSP- 366

Qy    694 PTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 753
          | | |     | :      |  ||   ||   |  |  |  |  | | || |:|
Db    367 ----KSPPPSPSPPPPYYVYKSPPPSPSPPPPYYYKSPPPSPSPPPPYYKSPPPSP 422

Qy    754 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKEPAPTTPKKPAPTTPETPP 811
          |  |        |  | :    |  | | | |||     | |  |  |:|
Db    423 SPP------PPYVYKSPPPSPSPPPPYYYKSPPPSPSPPPPYVYKSPPPSPSPPPPYYKSPPPSPSPPPP 475
```

Search completed: April 26, 2002, 16:31:27
Job time: 682 sec

```
                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:          April 26, 2002, 16:32:03 ; Search time 62.75 seconds
                                            (without alignments)
                                            796.401 Million cell updates/sec

Title:           AA2
Perfect score:   7276
Sequence:        1 MAWKTLPIYLLLLLSVFVIQ........ARAITTRSGQTLSKVVWNCP 1363

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters:    100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1175	16.1	5179	1	MUC2_HUMAN	Q02817 homo sapien
2	950	13.1	1664	1	SLP1_CLOTM	006852 clostridium
3	809	11.1	1367	1	AMYH_YEAST	P08640 saccharomyc
4	651	8.9	1255	1	MUC1_HUMAN	P15941 h mucin 1 p
5	559	7.7	875	1	FP1_MYTED	025460 mytilus edu
6	555.5	7.7	2700	1	ZAN_HUMAN	Q9y493 homo sapien
7	551	7.6	620	1	EXTN_TOBAC	P13983 nicotiana t
8	551	7.6	1087	1	NFH_MOUSE	P19246 mus musculu
9	530.5	7.3	1162	1	TCNA_TRYCR	P23253 trypanosoma
10	530	7.3	865	1	CPN_DROME	Q02910 drosophila
11	518.5	7.1	872	1	FP1_MYTCO	Q25434 mytilus cor
12	517.5	7.1	831	1	NFH_RAT	P16884 rattus norv
13	509.5	7.0	662	1	MUC1_XENLA	Q05049 xenopus lae
14	503.5	6.9	1970	1	RPB1_HUMAN	P24928 homo sapien
15	502	6.9	1020	1	NFH_HUMAN	P12036 homo sapien
16	498.5	6.9	1970	1	RPB1_MOUSE	P08775 mus musculu
17	493.5	6.8	467	1	RPB1_CRIGR	P11414 cricetulus
18	490	6.7	2142	1	BAT2_HUMAN	P48634 homo sapien
19	488.5	6.7	826	1	SSP2_PLAYO	001443 plasmodium
20	475.5	6.5	267	1	EXTN_MAIZE	P14918 zea mays (m
21	471.5	6.5	5376	1	ZAN_MOUSE	088799 mus musculu
22	468.5	6.4	634	1	HWP1_CANAL	P46593 candida alb
23	467	6.4	817	1	VRP1_YEAST	P37370 saccharomyc
24	454	6.2	797	1	VGLX_HSVEB	P28968 equine herp
25	454	6.2	1161	1	YJ9P_YEAST	P47179 saccharomyc
26	453.5	6.2	1794	1	YAV1_SCHPO	010172 schizosacch
27	448.5	6.2	670	1	VG50_HSVI1	Q00130 ictalurid h
28	446	6.1	1229	1	N121_HUMAN	Q9y2n3 homo sapien
29	444.5	6.1	2774	1	MAPA_RAT	P34926 rattus norv
30	442.5	6.1	751	1	FP1_MYTGA	Q27409 mytilus gal
31	442	6.1	1083	1	T2D2_HUMAN	002268 homo sapien
32	436.5	6.0	1411	1	TCOF_HUMAN	Q13428 homo sapien
33	434.5	6.0	3256	1	KI67_HUMAN	P46013 homo sapien
34	433.5	6.0	3164	1	TEGU_HSV11	P10220 herpes simp
35	432.5	6.0	439	1	XP2_XENLA	P17437 xenopus lae
36	428	5.9	2517	1	NCR2_HUMAN	Q9y618 h nuclear r
37	427.5	5.9	1251	1	YQU3_CAEEL	Q09550 caenorhabdi
38	426	5.9	2715	1	TRX2_HUMAN	Q9umn6 homo sapien
39	425.5	5.9	1185	1	DRPL_HUMAN	P54259 homo sapien
40	424.5	5.8	1125	1	MAP4_MOUSE	P27546 mus musculu
41	424	5.8	307	1	SGS3_DROME	P02840 drosophila
42	423.5	5.8	3421	1	TEGU_HSVEB	P28955 equine herp
43	422.5	5.8	1183	1	DRPL_RAT	P54258 rattus norv
44	419.5	5.8	2476	1	ZAN_PIG	Q28983 sus scrofa
45	411.5	5.7	907	1	VGP3_EBV	P03200 epstein-bar

ALIGNMENTS

RESULT 1
MUC2_HUMAN

```
ID   MUC2_HUMAN              STANDARD;      PRT;   5179 AA.
AC   Q02817; Q14878;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
GN   MUC2 OR SMUC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=94132002; PubMed=8300571;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT   "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT   Identification of the amino terminus and overall sequence similarity
RT   to prepro-von Willebrand factor.";
RL   J. Biol. Chem. 269:2440-2446(1994).
RN   [2]
RP   SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=93016075; PubMed=1404049;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA   Kim Y.S.;
RT   "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT   both upstream and downstream of its central repetitive region.";
RL   J. Biol. Chem. 267:21375-21383(1992).
RN   [3]
RP   SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX   MEDLINE=91358717; PubMed=1885763;
RA   Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA   Petersen G.M., Kim Y.S.;
RT   "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT   and polymorphism.";
RL   J. Clin. Invest. 88:1005-1013(1991).
CC   -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC       OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC       PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC       AGENTS AT MUCOSAL SURFACES.
CC   -!- SUBUNIT: MULTIMERIC.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC       BRONCHUS, CERVIX AND GALL BLADDER.
CC   -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC       INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC       VARIES AMONG DIFFERENT ALLELES.
CC   -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC       OF SILKWORM HEMOCYTIN.
CC   -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ----------------------------------------------------------------------
```

```
DR   EMBL; L21998; AAB95295.1; -.
DR   EMBL; M74027; AAA59875.1; -.
DR   EMBL; M94131; AAA59163.1; -.
DR   EMBL; M94132; AAA59164.1; -.
DR   MIM; 158370; -.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002400; GF_cysknot.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF00094; vwd; 4.
DR   PRINTS; PR00438; GPCYSKNOT.
DR   SMART; SM00214; VWC; 2.
DR   SMART; SM00216; VWD; 1.
DR   SMART; SM00011; VWD; 4.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01185; UNKNOWN_1.
DR   PROSITE; PS01186; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 2.
KW   Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21   5179
FT   DOMAIN     1401   1747       POTENTIAL.
FT   DOMAIN     1401   1416       MUCIN 2.
FT   REPEAT     1417   1432       APPROXIMATE REPEATS.
FT   REPEAT     1433   1448       1.
FT   REPEAT     1449   1464       2.
FT   REPEAT     1465   1471       3.
FT   REPEAT     1472   1478       4.
FT   REPEAT     1479   1494       5.
FT   REPEAT     1495   1517       6.
FT   REPEAT     1518   1533       7A.
FT   REPEAT     1534   1556       7B.
FT   REPEAT     1557   1572       8A.
FT   REPEAT     1573   1596       8B.
FT   REPEAT     1597   1612       9A.
FT   REPEAT     1613   1635       9B.
FT   REPEAT     1636   1651       10A.
FT   REPEAT     1652   1675       10B.
FT   REPEAT     1676   1683       11A.
FT   REPEAT     1684   1699       11B.
FT   REPEAT     1700   1715       12.
FT   REPEAT     1716   1731       13.
FT   REPEAT     1732   1747       14.
FT   REPEAT     ....   ....       15.
FT   REPEAT     ....   ....       16.
FT   DOMAIN     4815   4886       VWFC 1.
FT   DOMAIN     4924   4991       VWFC 2.
FT   DOMAIN     5075   5160       CTCK.
FT   DISULFID   5075   5122       BY SIMILARITY.
FT   DISULFID   5089   5136       BY SIMILARITY.
FT   DISULFID   5098   5152       BY SIMILARITY.
FT   DISULFID   5102   5154       BY SIMILARITY.
FT   DISULFID      ?   5159       BY SIMILARITY.
FT   CARBOHYD    163    163       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD    670    670       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD    770    770       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD    894    894       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   1139   1139       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   1154   1154       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   1215   1215       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   1230   1230       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   1246   1246       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   1787   1787       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   1820   1820       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4339   4339       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4351   4351       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4362   4362       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4373   4373       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4422   4422       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4438   4438       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4502   4502       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4616   4616       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4627   4627       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4752   4752       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4787   4787       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4881   4881       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4888   4888       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4955   4955       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   4970   4970       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   5019   5019       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   5038   5038       N-LINKED (GLCNAC) (POTENTIAL).
FT   CARBOHYD   5069   5069       N-LINKED (GLCNAC) (POTENTIAL).
FT   CONFLICT   1351   1351       H -> L (IN REF. 3).
FT   CONFLICT   1412   1412       T -> S (IN REF. 3).
FT   CONFLICT   1449   1449       L -> P (IN REF. 3).
FT   CONFLICT   1504   1504       M -> T (IN REF. 3).
FT   CONFLICT   4192   4192       G -> S (IN REF. 2).
SQ   SEQUENCE   5179 AA;  540295 MW;  85CD7571FB9A5663 CRC64;
```

```
Query Match            16.1%;  Score 1175;  DB 1;  Length 5179;
Best Local Similarity  29.3%;  Pred. No. 2.7e-38;
Matches  398;  Conservative  81;  Mismatches 519;  Indels 360;  Gaps 52;

Qy     36 SFERGRECDC----------DAQCKKYDKC----------CPDYESFCAEVHNPTSPSSKKAP 79
             :|:|           ||        :|:||
Db   1081 SCDTGGDCECFCSAVASYAQECTKEGACVPWRTPDLCP--IFC-DYYNPPHECEWHYEP 1136

Qy     80 PPSGASQTIKSTT--------------KRSPK--PPNKKKTKKVVIESEE-----ITEEH- 117
Db   1137 CGNRSFETCRTINGIHSNISVSYLEGCYPRCPKDRPIYEEDLKKCVTADKCGCYVEDTHY 1196

Qy    118 ---SVSENQESSSSSSSSSSSSSSSTI------------WKIKSSKNSAANRELQK 155
             |:|                  :|||
Db   1197 PPGASVPTEETCKSCVCTNSSQVCRPEEGKILNQTQDGAFCYWEI----CGPNGTVEK 1251

Qy    156 ---------KLKVKDNKKNRTKKKPTKRPPVV---- 178
Db   1252 HFNICSTTTRPSTLTTFTTTILPTTPSFTTTTTTTPTSSTVLSTTTPKLCCLWSDWINE 1311

Qy    179 DEAGSGLDNGDFK----------VTTPDTSTTQH-NKVSTSPK--------- 210
Db   1312 DHPSSSGSDDGDREPFPGVCGAPEDIECRSVKDPHLSLEQHQOKVQCDVSVGFICKNEDQF 1371

Qy    211 -----------IITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETK 246
Db   1372 GNGPFGLCYDYKIRVNCCWPMDKCITTPSPPTTTPSPPPTTTLPPTTPSPPTT---- 1427

Qy    247 ETTTTNKGTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPIPKAEHTTKGPALTTP 306
Db   1428 -TTTTPPTTTPSPPITTTTTTTP----LPTT----TPSPPISTTTTPPTTTTP 1470

Qy    307 KEPT-----PTTPKEPASTTPKEPTPTTTKEPAPTTKSAP-TTPKEP----APTTKSAP--TTP 352
Db   1471 SPPTTTPSPPTTTTPPTTTTTTPPPTTTTTTTPSPPTTTTKEPAPTTKEPAPTTPKKPA 1530

Qy    353 KEPAPTTTKEPAPTTPKEP----------APTTTKEPAPTTKSAP--TTPKEPAPTTPKKPA 403
Db   1531 TTPPPTTTPSPPTTTPITPPPTSTTTLPPTTTPSPPTTTTTTPSPPTTTTPSPPT 1590

Qy    404 PTTPKEPAPTTPKEPTPTTPKEPAPTTIKEPAPTTPKEP-----APTAPKKPAPTTPKE 456
Db   1591 ITTTTPPTTTTPPTTTTPSPPTTTKEPAPTTTLPPTTTPSPPTTTT 1650

Qy    457 PAPTTPKEPAPTTTKEPSPTTPKEPAPTTKSAP---TTPKEPSP 513
Db   1651 PPPTTTPSPPTTTTPSPPITTTTTTTPSSPITTTTTTPMTTPSPTTTPSPI 1709
```

```
Qy    514 TTTKEPAPTTPKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTP-              572
          ||||  |   ||   |   || :| | |:|    :||  || ||
Db   1710 TTTTPSSTTTPSPPPTTMTTPSPTTTPPSPPTTTMTTLPPTTPSSPLTTTPLPPSITPPT 1769

Qy    573 ---KETAPTTP---                                               585
             ||||
Db   1770 FSPFSTTTPPTTPCVPLCNWTGWLDSGRPNFHKRPGGDTELIGDVCGPGWAANISCRATMYP 1829

Qy    586 -----TTPEKLAPTT                                              595
               |:  ||
Db   1830 DVPIGQLGQTVVCDSVSGLICKNEDQKPFGGVIPMAFCLNYEINVQCCECVTQPTTMTTTTT 1889

Qy    596 PEKPAPTTPPEELAPTTPEEPTPT--TPEEPAPTTPKAAAPNTPKEPAPTTPKEP----AP 649
          | |  |  |||| | |  || |:    |  ||
Db   1890 TENPTPPTTPITTTTTVTPTPTGTQTPTPITTTTTTVTPTPTGTQTPTPITTTT       1949

Qy    650 TTPKEPAPT-TPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKEPTSTTSDKP   708
            | | |     ||  | | | | |   |  ||  | :|   |  |  |  |  :
Db   1950 TTTTVTPTPTPGTQTPTPITTTTTTVTPTPTGTQTPTPITTTTTTVTPTPTGTQ       2009

Qy    709 AP-TTPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTK         767
           |  | | | |  ||  |   || | | |  |  || | :|   |  |  |
Db   2010 TPTTTPITTTTTVTPTPTPGTQTPI-TTPITTTTTTVTPTPTGTQTPTPITTTTTTTV    2068

Qy    768 GPTSTTSDKPAP--TTPKETAPTTPKEPAPTTPKKPAPTTPEPPTTSEVSTPTTTPKEPT  826
              |   | |    ||  | | || | | |  || | |  || |  || :| |  || |
Db   2069 TPTPTPGTQTPTPITTTTTTVTPTPTGTQTPTPITTTTTTTVTPTPTGTQTPT        2127

Qy    827 TIHKSPDESTPELSAEPTPKALENSKEP------                           875
           | | |  ||    |  ||   | | |
Db   2010 ----------GVPT-IKTPAATKPEMTTAK                               875

Qy    876 DKTTERDLR---TTP--ETTTAAPKMT---KETATTTEKTTESKITAITTQVTSTTTQDT  927
           |   |     ||    |  | |       ||   ||  ||  | | || : | |  |
Db   2128 T---TPITTTTTVTPTPTPGTQTPTPITTTTTTVTPTPITGTQTPTPITTTTTTVI      2184

Qy    928 TPFKITTTLKTTTLAPKVTT--KKTITTTEIMNKPEETAKPKDRATNSKATTPKPOKPTKA 986
           || | ||    |  | |      | |  |    || |  | |  |  | | | |  |:|
Db   2185 PTPTPGTQTPTPTTTTTTVTPTPITTTTTTTVTPTPTGTQTPTPTP-TGTQTPTPTT    2243

Qy    987 P-KKPSTKKPKPKTMPRVRKPKTTPPPRKMTSTMPELNP---                   1041
          |  | | | || | |     | ||  |:    |  |:
Db   2244 TP--ITT--TTTVTPTPTPGTQTPTPITTTTTTVTPTPTPGTQTPTPITTTTTTTVI    2299

Qy   1042 PNSKLVENPKSEDAGGAEGETPHMLLRPHVFMPEVTP                        1079
          | |           |    || | | |    | | |
Db   2300 PTPTPTPGTQTPTPITTTTTTVTPTPTGTQTPTPITTTTTTTVTPTP-TSRIAEAMLQTTTR-PNQT 2359

Db   2360 PITTTTTVTPTPTPG---TQTP-TTTPITTTTTTVTP                        2392
```

```
DR   EMBL; X67506; CAA47841.1; -.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 3.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   Cell wall; S-layer; Signal; Repeat.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   1664       CELL SURFACE GLYCOPROTEIN 1.
FT   DOMAIN       36    763       4 X 156 AA APPROXIMATE REPEATS.
FT   REPEAT       36    191       1.
FT   REPEAT      207    363       2.
FT   REPEAT      409    565       3.
FT   REPEAT      607    763       4.
FT   DOMAIN      771   1377       APPROXIMATE TANDEM REPEATS OF
FT   DOMAIN     1378   1449       T-P-S-D-E-P.
FT   DOMAIN     1453   1494       GLY/PRO/SER/THR-RICH.
FT   DOMAIN     1495   1565       SLH 1 (INCOMPLETE).
FT   DOMAIN     1566   1625       SLH 2.
FT   DOMAIN     1566   1646       SLH 3.
FT   DOMAIN     1626              SLH 4 (INCOMPLETE).
SQ   SEQUENCE 1664 AA;  178194 MW;  5F39669SBA9FE74B CRC64;
```

```
Query Match             13.1%;
Best Local Similarity   31.0%;
Matches   312;   Conservative   104;   Mismatches  361;   Indels  230;   Gaps  55;
```

```
Qy    286 VLAKPTP-KAETTTKGPALTTTPKEPTPTTTPKEPTPTTTPKEPASTTTPKEPTPTTIKSAPTTTKEPAPTT 344
             : | |          || |   || | |   ||
Db    758 VVIQPAPIKAASDEPIPTDTPSDEPTPS--------DEPTPS---DEPTPSDEPTPSD               804

Qy    345 TKSAPTTTPKEPAPTTTKEPAPTTPKEPA---PTTTKEPAPTTTKSAPTTPKEPAPTTPK            400
          | |  || |  |                   |
Db    805 EPTPSETPEEPIPIDTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTP                  864

Qy    401 KPAPTTPKEPAPTTPKEEPTPTTTPKEPAPTTKEPAPT-TPKEPAPTAPKKPAPTTPKEPAP          459
           | | |          | |
Db    865 SDEPTPSDEPTPS--DEPTP--DEPTP-SDEPPPSEPEEPTPTDTPSDEPTPSDEPTP              919

Qy    460 TTPKEPA--PTTTKEPSPT-TPKEPAPTTKS-APTTTKEPAPT----TTKSAPTTTKEPS            512
           | | |           |      |
Db    920 SDEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPEEPIPTDTPSDEPTPSDEPTPSDEPPTT            979

Qy    513 ----PTTTKEPAPT-TPKEPAPTTPKEPAPTTPKEPAPTTPKKPAPTTKRAPTAPKEP              567
              | | |       ||
Db    980 PSDEPTPSDEPTPSETPEEPIPIDTPSDEPTPSDEPTPSDEPTPSDEPTPSD                   1039

Qy    568 APT---TPKETAPTTPKKLTPTTTPEKLAPTTPEKPAPT---TPEELAPTTPEEPTPT-TP            620
          | |   | |
Db   1040 TPSDEPTPSE----TPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSETP                  1094

Qy    621 EEPAPTTPKAAAPNTPKEPAPTTPKEPA--PTTPKEPAPT-TPKETAPTTPKGTAPTTLK            677
          || | |         | |
Db   1095 EEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPSD          1154

Qy    678 EPAPTTPKKPAPKELAPTTTKEPT-STTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTT           736
          || | |
Db   1155 EPTPS--DEPTPSD-EPTPSDEPTPSETPEEPIPTDTPSDEPTPS--DEPTPS--              1208

Qy    737 PKGTAPTTLKEPAPT-TPKKPAPKELAPTTTKGPTSTTTSDKPAPT---TPKETAPTTPKE          792
             | | |           |
Db   1209 ---DEPTPSDEPTPSEPEEPI-----------PTDTPSDEPTPSDEPTPSDEP-EPTPSDE         1253
```

```
RESULT  2
SLP1_CLOTM
ID   SLP1_CLOTM     STANDARD;      PRT;   1664 AA.
AC   Q06852;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
DE   PROTEIN 1).
GN   OLPB.
OLB.
OS   Clostridium thermocellum.
OC   Bacteria; Firmicutes; Bacilli/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1515;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93209931; PubMed=8458832;
RA   Fujino T., Beguin P., Aubert J.-P.;
RT   "Organization of a Clostridium thermocellum gene cluster encoding the
RT   cellulosomal scaffolding protein CipA and a protein possibly involved
RT   in attachment of the cellulosome to the cell surface.";
RL   J. Bacteriol. 175:1891-1899(1993).
```

```
-!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
    GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
    WITH RELEASE OF BETA-D-GLUCOSE.
-!- SIMILARITY: TO S.POMBE SPBC215.13.
-!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
```

```
DR   EMBL; 238061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA33069.1; -.
DR   PIR; B26877; B26877.
DR   PIR; A26878; A26877.
DR   PIR; S48478; S48478.
DR   SGD; S0001458; MUC1.
DR   SGD; S00014578; MUC1.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL         1     21
FT   CHAIN         22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN       210   1367       SER/THR-RICH.
FT   CARBOHYD     817    817       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     874    874       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;

Query Match             11.1%;  Score 809;  DB 1;  Length 1367;
Best Local Similarity   27.5%;  Pred. No. 9.8e-25;
Matches  328;  Conservative 112;  Mismatches 526;  Indels 226;  Gaps 53;
```

```
Qy    44  DCDAQCKKYDK---CCPDYESFCAEVHNPTSPPSSKKAPPSGASQTIKSTTIKRSPKPPN  100
          ||                                                  ||
Db   174  DLSTGCNNYDNQGHSQTDFPGFYWNIDCDNNCGGTKSSTTTSSTTESS--STT-------  223

Qy   101  KKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVK  160
Db   224  ---TSSTSESSTTTSSTSESSTTTSSTSESSTTTSSSTT-----APATPTTTSCTTEKPTP  274

Qy   161  DNKKNKKKKKPTPKPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPR  220
Db   275  PTTTSCTKEKPT--PPHHD--------------------TTPCT----KKKTTSKTCTTKKTTTPV  314

Qy   221  PSLPPNSDTSKETSLIVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDL  280
Db   315  PT--PSSSTTESSS------APVPTPSSSTT------ESSSAPVTSSTTESSS---  353

Qy   281  APTSKVLAKPTPKAEITTKGPALTTPKEPTTTPKEPASTTPKEPTPTTIKSAPTTPKEP  340
Db   354  APV-------PTPSSSTTESSSAPVTSSTTESSSAPVPTSSSTTESSSAPVPT---SSTTESSSAPV  407

Qy   341  APTT-----TKSAP-----TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAP--TTTKEPAPTTKSA  387
Db   408  APVTSSTTESSSAPVTSSTTESSSAPVTSS----TTESSSAPVTSSTTESSSAPVTSS  463

Qy   388  PTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTP---TTPKEPAPTTKEPAPTTPKEP  444
Db   464  STTESSSAPVT---SSTTESSSAPV----PTPSSSTTESSSAPVT---SSTTESSSAPV  512

Qy   445  APKKPAPTTPKEPAPTTKEPAPTTTKEPAPTTPKEP-SPTTTPKEPAPTTTKSAPTTTKEPAPTTTKSA  504
Db   513  ------PTPSSSTTESSSAPATPS---SSTTESSSAPVTSS---TTESSSAPVTPSS  559

Qy   505  PTTPKEPSPTTTMKEPAPTTPKEPAPT-TPKKPAPTTPKEPAPTTPKEPAPTTKKPAPTA  563
Db   560  STTESSSSPVTSS---TTESSSAPVPTPSSSTTESSSAPVPT---PSSSTTESSSAPA  611

Qy   564  PKEPAPTTPKETAPTTPKKLTPTTPEKLIAPTTPEKPAPTTPEELAPT-TPEEPTPTTPEE  622
```

```
Qy   793  PAPTTPKKPAPT--TPETPPPTTSEVSTPTTIKEPTTIHKSPDESTPELSAEPTPKALENS  851
Db   1254 PTPS--DEPTPSEPEEPIPTDIPDSPEPTPSDEPT----PSDEPTP--SDEPTPSDEPTP  1305

Qy   852  PKEPGVPTTKTPAAT--KPEMTTTAKDKTTERDLRTTPEPTTTRAAPKMTKETATTTEKTTE  909
Db   1306 SDEPTPSDEPTPSETPEEEPIPTDIPDSPEPTPSD----EPTPSDEPTPSDEPTPSDEPTPS  1361

Qy   910  SKIPTATTTQVTSTTTQDTTPTFKITTLKTTTLAPKVTTTKKTTITTTEIMNKPEETAKPKDR  969
Db   1362 DEPTPSDEPTPSETPEEPTP-----TTTPTTPTSTI  1396

Qy   970  ATNSKAT---------TPKPQKPTKAP-KKPTSTKKPKTMPRVRKPKTTPTPTPRKM------  1014
Db   1397 SGGSGSGGGGGGGGGGGGVPTPSPTPTTPTSKPTSTPAP----TEIEEPTPSDVPGAIGGEHRA  1453

Qy   1015 -----TSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLIVEVNPKSEDAGGAEGETPHM-----  1066
Db   1454 YLRGYPDGSFRPERNITRAEAAVIP-------AKLL----GADESYGAQSASPYSD  1498

Qy   1067 LLRPH-----------------VFMPE------------VTPDMDYLPRVPNQGI--------  1092
Db   1499 LADTHWAAWAIKFATSQGLFKGYPDGTFKPDQNITRAEFATVVLHFLTKVKGQEIMSKLA  1558

Qy   1093 ----IINPMLSDETNICNG------KPVDGLTTL------RNGTLVAPRGHYFWMLSPFSPPS  1139
Db   1559 TIDISNPKFDD------CVGHWAQEFIEKLISLGYISGYPDGT--------FKPQN  1600

Qy   1140 PARRITEEVWGIPSPIDTVFFTRCNCEGKTFFFKD--SQYWRFTNDIKD  1184
Db   1601 YIKRSESV---ALINRALERGPLNGAAPKLFPDVNESYWAF-GDIMD  1642
```

```
RESULT   3
AMYH_YEAST
ID   AMYH_YEAST          STANDARD;    PRT;   1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
DE   GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=5288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
```

```
Db   612 PTPSSSTTESSSAPVT----SSTTESSSAP-VPTPSSSTTESSSAPVPTPSSSTTESSSA 666
             ||:|      |||                 |||
Qy   623 PAP------TTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTT-TPKGTAPTTL 676
             ||:|      |||
Db   667 PVPTPSSSTTESSSAPVT----SSTTESSSAPVT---SSTTESSSAPVPTPSSSTTESS 718
             ||:|      |||:|            ||:|      |||
Qy   677 KEPAPTTPKKPAPKELAPTTPKEPTSTTSDKPAPTTPKGTAPTTPKEPAP---TTPKEPA 733
             ||:|      |||:|            ||:|      |||
Db   719 SAPVPTPSSSTTESSSAPVPTPS-SSTTESSSAPVTSSSTTESSSAPVTPSSSTTESSSA 777
             ||:|      ||:|            ||:|      |||
Qy   734 PT-TPKGTAPTTLKEPAPTTPKKPAPKELAPTTKGPTSTTSDKPAPTTPKETAPTTPKE 792
             ||:|      |||          |||          |||
Db   778 PVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPTPPSSSTTESSSV 837
             |||          |||
Qy   793 PAP------TTPKPAPTTPEPPPPTTPSEVSTPT-----TTKEPTTIHKSPDESTPELSAE 842
             :||          |:||          ||:|
Db   838 PVPTPSSSTTESSSAPVPSSSTTESSSAPVPTPSSSSNTTSSAPSSI--PPSSTTESFST 894
             :||          |:||          ||:|
Qy   843 PTPKALENSPKEPGVPTTKKPAATKPEMTTTAAKDKTTERDLRTTPDTTTTAAPKMTKETAT 902
             |||          |||          ||:|
Db   895 GT-TVTPSSSKYPGSQTETSVSSTIETTIVPIKTTTVSVTTPSTTTITTVCSTGTNSAGE 953
             |||          |||
Qy   903 TTEKTTESKIITAT--TTQVSTTTDQTTTPFKITTLKTTILAPKVTT---TKKIITTT-EI 956
             |||          |:||          ||:|
Db   954 TTSGGCSPKTVTTTVPTTTTISVTTSSTTTTTTVCSTGTNSAGETTSGGCSPKTITTTTVPC 1013
             |||          |:||          ||:|
Qy   957 MNKPEETAKPKDRAINSKAATTPKPQKP----------TKAPKKPISTKKPKTM 999
             |:|          ||:|
Db  1014 STSPSETA-----------SESTTTSPITPVATTVVSTTVVTTEYSTSTKPGGEITTIFVTKNI 1065
             |:|          ||:|
Qy  1000 PRVRKPKTTPTPRKMTMSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGA 1059
             ||:|          |:|
Db  1066 PTTYLTTIAPTP--SVTTTVPTTTT-IITTVCSTGT------------NSAGETTSGC 1109
             ||:|          |:|
Qy  1060 EGETPHMLLRPHVHFMPEVTPDMDYLPRVPNQGIIINPMLSDETNICNGKPVDGLITLRNG 1119
             |||          :||
Db  1110 SPKTVTTTVPCSTGTGEYTTEATLVTTAVTTVVTTESSTGTNSA-GKTTTGYTTIKSVP 1168
             |||          :||
Qy  1120 TLVAFRGHVFWMLSPFSPPPSPARITEVWGIPSPIDTVTPTRC-----NCEGKT 1167
             :||          |:|
Db  1169 TT------------YVTTLAPSAPVTPAIN----AVPTITT--TECSAATNAAGET 1207
             :||          |:|
Qy  1169 TT------------YVTTLAPSAPVTPAIN----AVPTITT--TECSAATNAAGET 1207
```

RESULT   4

ID   MUC1_HUMAN   STANDARD;   PRT;   1255 AA.
AC   P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM)
DE   (EPISIALIN) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE   (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-
DE   REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE   DF3).
GN   MUC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=90368716; PubMed=2394722;
RA   Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RL   "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL   J. Biol. Chem. 265:15294-15299(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90202794; PubMed=2318825;
RA   Ligtenberg M.J., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT   "Episialin, a carcinoma-associated mucin, is generated by a
RT   polymorphic gene encoding splice variants with alternative amino
RT   termini.";
RL   J. Biol. Chem. 265:5573-5578(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90368715; PubMed=1697589;
RA   Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA   Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT   "Molecular cloning and expression of human tumor-associated
RT   polymorphic epithelial mucin.";
RL   J. Biol. Chem. 265:15286-15293(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=91097524; PubMed=2268309;
RA   Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA   Taylor-Papadimitriou J., Gendler S.J.;
RT   "Structure and expression of the human polymorphic epithelial mucin
RT   gene: an expressed VNTR unit.";
RL   Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90276413; PubMed=2351132;
RA   Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
RA   Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA   Keydar I.;
RT   "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT   may generate multiple protein forms.";
RL   Eur. J. Biochem. 189:463-473(1990).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=90276414; PubMed=2112460;
RA   Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horov J.,
RA   Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT   "A transcribed gene, containing a variable number of tandem repeats,
RT   codes for a human epithelial tumor antigen. cDNA cloning, expression
RT   of the transfected gene and over-expression in breast cancer
RT   tissue.";
RL   Eur. J. Biochem. 189:475-486(1990).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91033045; PubMed=1688329;
RA   Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA   Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT   "Isolation and characterization of an expressed hypervariable gene
RT   coding for a breast-cancer-associated antigen.";
RL   Gene 93:313-318(1990).
RN   [8]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=88330762; PubMed=3417635;
RA   Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA   Burchell J.;
RT   "A highly immunogenic region of a human polymorphic epithelial mucin
RT   expressed by carcinomas is made up of tandem repeats.";
RL   J. Biol. Chem. 263:12820-12823(1988).
RN   [9]
RP   SEQUENCE OF 1-169 FROM N.A.
RX   MEDLINE=90088473; PubMed=2597151;
RA   Abe M., Siddiqui J., Kufe D.;
RT   "Sequence analysis of the 5' region of the human DF3 breast
RT   carcinoma-associated antigen gene.";
RL   Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN   [10]
RP   SEQUENCE OF 1-109 FROM N.A.
RC   TISSUE=Thyroid;
RX   MEDLINE=96183746; PubMed=8608966;
RA   Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT   "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT   transcriptase polymerase chain reaction of the MUC1 gene.";
RL   Int. J. Cancer 66:55-59(1996).
RN   [11]

```
RP   SEQUENCE OF 1-89 FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=96181716; PubMed=8604237;
RA   Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA   Lee L.N., Luh K.T., Wu C.W.;
RT   "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT   tissues.";
RL   Oncology 53:118-126(1996).
RN   [12]
RP   SEQUENCE OF 1-46 FROM N.A.
RC   TISSUE=Breast carcinoma;
RA   Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
RA   Submitted (OCT-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC       CYTOSKELETON.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
CC       IS ALSO PRODUCED.
CC   -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
CC       TUMORS, SUCH AS BREAST CANCER.
CC   -!- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND SIALIC
CC       ACID).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT
CC       VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
CC       MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J05582; AAA60019.1; -.
DR   EMBL; M32738; AAA35804.1; -.
DR   EMBL; M32739; AAA35806.1; -.
DR   EMBL; J05581; AAA59876.1; -.
DR   EMBL; M61170; AAB53150.1; -.
DR   EMBL; X52229; CAA36478.1; ALT_SEQ.
DR   EMBL; X52228; CAA36477.1; ALT_SEQ.
DR   EMBL; M35093; AAB59612.1; ALT_SEQ.
DR   EMBL; Z17324; CAA78972.1; -.
DR   EMBL; Z17325; CAA78973.1; -.
DR   EMBL; M31823; AAA35757.1; -.
DR   EMBL; S81781; AAD14376.1; ALT_INIT.
DR   EMBL; S81736; AAD14369.1; ALT_INIT.
DR   EMBL; M21868; AAA59874.1; ALT_SEQ.
DR   PIR; A35175; A35175.
DR   PIR; B35175; B35175.
DR   PIR; S10218; S10218.
DR   GlycoSuiteDB; P15941; -.
DR   MIM; 158340; -.
DR   MIM; 113720; -.
DR   InterPro; IPR000082; SEA.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS50024; SEA; 1.
DR   PROSITE; PS00024; SEA; 1.
KW   Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW   Repeat; Alternative splicing.
FT   SIGNAL        1     23
FT   CHAIN        24   1255
FT   DOMAIN       24   1162       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1163   1186       POTENTIAL.
FT   DOMAIN     1187   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       81    960       44 X 20 AA TANDEM REPEATS.
FT   DOMAIN     1034   1151       SEA.
FT   CARBOHYD    957    957       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    975    975       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1029   1029       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1055   1055       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1077   1087       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   1088   1255       MISSING (IN SECRETED ISOFORM).
FT   CONFLICT    134    134       P -> Q (IN REF. 9).
FT   CONFLICT    154    154       P -> Q (IN REF. 9).
FT   CONFLICT   1021   1021       S -> Y (IN REF. 3).
FT   CONFLICT   1251   1251       A -> T (IN REF. 3).
SQ   SEQUENCE   1255 AA;  122072 MW;  5E28DFC4DE7D9A82 CRC64;
```

```
Query Match              8.9%;        Score 651;  DB 1;   Length 1255;
Best Local Similarity   27.8%;        Pred. No. 1e-18;
Matches  295;  Conservative  72;  Mismatches  460;  Indels  234;  Gaps  51;

Qy    233 TSLIVNKETTVEIKETTTTTNKQTSTDGKEKTTSAKETQSIEKTTSAKD-LAPTSKVLAKPT 291
              |||:  :  ::  |::|||:|| ::|:||: |::: |:|: | |:  |:   :  :
Db     16 TVLIV----------VTGSGHASSTPGGEKEISATQRSSVPSSTEKNAVSMTSSVLSSHS 65

Qy    292 P-KAETTTKGP-ALTTTKEP-----------------TPTTTPKEPASTTPTPKEPTPPTTI---K 331
            |  :: ||  |:||||           |||   |||:| :
Db     66 PGSGSSTTQGSGDVTLAPATEPASGSAATWGQODVTSVSVPVTRPALGSTTPPAHDVTSAPDNK 125

Qy    386 SAPTTPKEPAPTTP------------------KKPAPTTPKEP--------APTTPKEPTPTP---- 423
            |||  |||       |||  :  ||
Db    180 SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTI 239

Qy    332 SAPTTPKEPAPTTPTIKSAPTTPKEPAPTTPKEP------APTTPKEPAPTTPKEPAPTTK 385
            |||  |||   :  :|| : |||    |||:|  |||   :  |:|
Db    126 PAPGSTAPPAHGVT-SAPDT--RPAPGSTAPPAHGVTSAPDT--RPAPGSTAPPAHGVT- 179

Qy    424 ----KEPAPTTKEP------------APTTPKEPAPTAPK----------KPAPTTPKEPAPTTPK---E 464
               ||||    :|| : ::|  ||
Db    240 SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT--RPAPGSTAPPAHGVT 299

Qy    465 PAPTTTKEPSPTTTPKEPAPTT---------TKSAPTTTKEPAPTTPKEPSPTTTKEP--- 519
             |||  ||| :||  |||    |:  ||| :  :|| |||
Db    300 SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT--SAPDT--RPAPGSTAPPAH 356

Qy    520 ----APTTPKEPAPTTPK---KPAPTTPKEPAPTTPK------EPAPTTTKKPAPTAPKEP 567
                |||  :|||     |||  |||    |  : |||
Db    357 GVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAH 416

Qy    568 APTTPKETAPTTPKLLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTT 627
             |||     |||     ||| :  ||| :::  ||  |:  ||  |  || :|
Db    417 GVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTA-----PAHGVTSAPDTRPAPGSTA 472

Qy    628 PKA------AAPNTPKEPAPTTPK-----EPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 679
            |       : :|| :  |:|||   |:| : |::|  |: :: || :  ||
Db    473 PPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP----PAHGVTSAPDTRPAP 528

Qy    680 APTTP-------KKPAPKELAPTTTKEPTSTISDKPAP--TTPKGTAPTTPKEPAPT 727
             |||     ||| :||||    :   :|| |:|   ||
Db    529 GSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP-----PAHGVTS 583

Qy    728 TPKEPAPTTPKG-----TAPTTLKEPAPTTP--------KKPAPKELAPTTTKGPTSTT 773
             ||||     : :|   :||| :  :   :|| |||:||  |  :|| :| ::
Db    584 TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP----PAHGVTSAP--APD 642

Qy    774 SDKPAP--TTPK-----ETAPTTPKEPAPTTP---------KKPAPTTPETPPPTTSEVS 817
            | |||    ||| :||||    :|||    |:|| :::|||||:||
Db    643 DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGS--TAPPAHGVTS 700

Qy    818 TPTTTKEPTT-----IHKSPDESTPELLSAEPTPKALENSPKEPGVPTTKTPAAITKPEM 870
            ||| |||     : : ::|    ||:  |  |: |: : |
Db    701 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAP-----DTRPAPGSTAP-- 753
```

RESULT 5

```
ID   FP1_MYTED     STANDARD;      PRT;   875 AA.
AC   Q25460;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
DE   PROTEIN 1) (MEFP1) (FRAGMENT).
GN   FP1.
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=6550;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91025829; PubMed=1367451;
RA   Filpula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
RT   "Structural and functional repetition in a marine mussel adhesive
RT   protein.";
RL   Biotechnol. Prog. 6:171-177(1990).
RN   [2]
RP   PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
RX   MEDLINE=83135732; PubMed=6298211;
RA   Waite J.H.;
RT   "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
RT   hydroxyproline-containing decapeptide in the adhesive protein of the
RT   mussel, Mytilus edulis L.";
RL   J. Biol. Chem. 258:2911-2915(1983).
CC   -!- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC       PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC       ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC       FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC   -!- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC   -!- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-K IS POST-TRANSLATIONALLY
CC       MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
CC       HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
CC       (DOPA) DERIVED FROM TYROSINE.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; X54422; CAA38294.1; -.
DR   InterPro; IPR002964; Adhesive_plaq.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01216; ADHESIVEI.
DR   PRINTS; PR01217; PRICHEXTENSN.
KW   Repeat; Hydroxylation.
FT   NON_TER        1      1
FT   NON_TER      870    870
FT   DOMAIN        67    870    TANDEM REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
SQ   SEQUENCE   875 AA;  100412 MW;  6EA85312748CAACE CRC64;
```

```
Query Match             7.7%;  Score 559;  DB 1;  Length 875;
Best Local Similarity  26.9%;  Pred. No. 2.6e-15;
Matches   281;  Conservative  102;  Mismatches  403;  Indels  260;  Gaps  56;
```

```
Qy    871 TTTAKDKTTERDLRTTPETT----------TAAPKMTKETATTTESKIATTTQVTSTT 923
              :       ::                    :     ::
Db    754 --PAHGVTSAPDTPRAPGSTAPPAHGVTSAP---DTRPAPGSTAPPAHGVTSAPDTRPA 807

Qy    924 TQDTTP--FKITTLKTTTLAPKVTTTKKTITTTT--EIMNKPEETAKPKDRATNSKATTPK 979
             : ::                :                ::
Db    808 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 867

Qy    980 PQKPTKAPKKPPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTTSRIAEAMLQTTTRPN 1039
             : :            :           :            ::
Db    868 P--GSTAPPAHGVTSAPDTRP---APGSTAPPAHGVTSAPDTRPA--------PG 909

Qy   1040 QT-PNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPML 1098
             :                          :                      :
Db    910 STAPPAHGVTSAPDTRPAPGSTAPPAH---------GVTSAPDNRPALGSTA---PPV 955

Qy   1099 SDETNICNGRKPVDGLITTLRNGTLVAFRGHYFWMLSPFSPPS 1139
              ::   :              ::                :::  ::
Db    956 HNVTSASGSASGSASTLVHNGTSARATTTPASKSTPFSIPS 996

Qy     67 HNPTSPPSSKKAPPSGASQTIKSTTKRSPKPNKKKTKKVIESSEEITEEHSVSENQESS 126
             : :                 :                ::
Db      3 HEPVYKP----------------VKITSYSAPYKPPPTYQPLKKKVD---------YRPT 35

Qy    127 SSSSSSSSSSTIWKIKSSKNSAANRELQKKLKVKDNKKNRTKKKPTKRPPVVDEAGSGLD 186
             ::            :                       :
Db     36 KSYPPTYGSKTNYLPLAKLSS------YKPIKTTYNAK--------TNVPPVY-------- 75

Qy    187 NGDFKVTTPDTSTTQHDNKVYSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETK 246
                     :       :       :               :
Db     76 ----------KPKMTYPPTYKPKPSYP----------------------TYKSKPTYKPK 104

Qy    247 ETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTP 306
                             :         :          :
Db    105 IT--------------YPPTYKAKPSYPSSYKPKKTYPPTYK-----PKLTY----PPTYKP 143

Qy    307 KEPTPTT---PKEPASTTPKEEPTPTTIKSAPTTP----KEPAPTTTKSAPTTPKEPAP 357
             : :        ::          :               :
Db    144 KPSYPPTYKPKPSYPPTYKTKITYPSSYKAKPSYPPTYKAKPSYP----PTYKAKPSYP 199

Qy    358 TTTKEPA-PTTTKEPAPTTTKEPAPTTTKSAPTTPKEPA--PTTPKKPA--PTTPKEPA 411
              : :         :          :
Db    200 TYKAKPTYKAKPTYPDSTYKAKPSYPPPTYKAKPTYKAKPSYPPTYKAKPSY 259

Qy    412 -PTTPKEPT-PTTPKEPAPTTPKEPAPTTPKEPA----PTTPKEPA--PTTPKEP 465
              : : :      : :          :
Db    260 PPTYKAKPTYKAKPTYKAKPTYKAKPSYP----PTYKAKPSYPPTYKAKP 315

Qy    466 A--PTTTKEPS--PTTPKEPA-PTTTKSAPTTPKEPA--PTTTKSAPTTPKEPSPTTTKEP 519
                 :                :             :
Db    316 SYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKPTYPSSYKAKPSYP----PTYKAKP 371

Qy    520 APTTPKEPAPTTPKKPA--PTTPKEPA--PTTPKEPAPTTTKKPAPTAPKEPAPTTPKET 575
              : :             :          :
Db    372 SYPPTYKAKPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKP-------T 424

Qy    576 APTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKRAAAPNT 635
              : :          :          :
Db    425 YPST-YKAKPSYP----PSYKAKPSYPPTYKAKPTY------------K 468

Qy    636 PKEPAPTTPKEPAPTTPKE----PAPTTPKETAPTT--PKGTAPTTLKEPAPTTPKKPAP 689
             :            :          :
Db    469 PSYPASYKAKPSYPPTYKSKSSYPSSYKPKKTYPPTYKPKLITYKPTYK-PKPSYPPSYKP 527

Qy    690 KELAPTTTKPEPTSTTSDKPAPTTPKGTAPTT-----PKEPAPTTPKEPAPTTPKGT--AP 742
             :            :          :
Db    528 KTTYPPTYK----------PKISYPPTYKAKPSYPATYKAKPSYPPTYKAKPSYPP 573

Qy    743 TTLKEPA--PTTPKKPA--PKELAPTTTKG-PTSTTSDKPAPTTPKETAPTTPKEPA--P 795
              : :         :          :
Db    574 TYKAKPSYPPTYKAKPSYKAKPTYPSTYKAKPSYPPTYKAKPSYP----PTYKAKPSYP 629

Qy    796 TTPKKPA-PTT---PETPPPTTSEVSTPTTKE----PTTIHKSPDESTPELSAEPTPK 846
             :           :          :
Db    630 TYKAKPTYPSTYKAKPSYPPTYKPRISYPPTYKAKPSYPPT-YKAKPSYPPTYKAKPTYK 688

Qy    847 ALENSPK------EPGVPTT--KTPA-----ATKPEMTTTARDKTTTERDLRTTTPETTTAAPK 895
             :             :         :
Db    689 AKPINPSTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKPSYPPTYKPSTYKAKPSYPPTYK 747
```

```
Qy   896 MTKETATTTEKTTESKITATTTQVTSTTTQDTTPFKITTLKTTTLAPKVTTTKKTITTTE 955
             |:||        ||  :||           |:|
Db   748 -TYKAKPTYPPTYKAK------PSYPPTYKPKPSYPPTYKSKSIYPSSYKPKKTYPPT- 798

Qy   956 IMNKPEETAKPKDRATNSKATTPKPQKPTKAPKK--PTSTKKPKIMPRVRKPKTIPTP- 1011
             :| |||                    ||
Db   799 --YKPKLTYPPTYK------PKPSYPPSYKPKITYPSTYKLKPSYPPTYKSKTSYPPT 848

Qy  1012 -----------RKMTSTMPELNPTSR 1026
                     |:||  :||:|
Db   849 YNKKISYPSSYKAKTSYPPAYKPTNR 874
```

RESULT   6

```
ID   ZAN_HUMAN      STANDARD;      PRT;   2700 AA.
AC   Q9Y493; O00218;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ZONADHESIN (FRAGMENT).
GN   ZAN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-2379 FROM N.A.
RX   MEDLINE=99018118; PubMed=9799793;
RA   Glockner G., Scherer S., Schattevoy R., Boright A., Weber J.,
RA   Tsui L.C., Rosenthal A.;
RT   "Large-scale sequencing of two regions in human chromosome 7q22:
RT   analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT   reveals 17 genes.";
RL   Genome Res. 8:1060-1073(1998).
RN   [2]
RP   SEQUENCE OF 2338-2700 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=97271566; PubMed=9126492;
RA   Gao Z., Harumi T., Garbers D.L.;
RT   "Chromosome localization of the mouse zonadhesin gene and the human
RT   zonadhesin gene (ZAN).";
RL   Genomics 41:119-122(1997).
CC   -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC       OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC       SIGNALING.
CC   -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC       APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC   -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC       ZONA PELLUCIDA.
CC   -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC       THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC       SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC   -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC       OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC   -!- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
```

```
DR   EMBL; AF053356; AAC78790.1; -.
DR   EMBL; U83191; AAC51208.1; -.
DR   MIM; 602372; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000998; MAM.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR002919; TIL.
DR   InterPro; IPR003328; TILa.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00629; MAM; 3.
DR   Pfam; PF01826; TIL; 4.
DR   Pfam; PF02345; TILa; 4.
DR   Pfam; PF00094; vwd; 4.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00137; MAM; 2.
DR   SMART; SM00214; VWC; 1.
DR   SMART; SM00011; VWC_def; 3.
DR   SMART; SM00216; VWD; 3.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS50060; MAM_2; 4.
KW   Glycoprotein; Transmembrane; Cell adhesion; Repeat.
FT   NON_TER         1      1
FT   DOMAIN       2293   2684       ? (MUCIN-LIKE DOMAIN).
FT                                  66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT   DOMAIN       1862   2292       VWFD 5.
FT   DOMAIN       1455   1861       VWFD 4.
FT   DOMAIN       1066   1454       VWFD 3.
FT   DOMAIN        953   1065       VWFD 2.
FT   DOMAIN        483    951       VWFD 1 (PARTIAL).
FT   DOMAIN        322    446       MAM 4.
FT   DOMAIN        161    326       MAM 3.
FT   DOMAIN        112    136       MAM 2 (PARTIAL).
FT   DOMAIN         <1    109       MAM 1.
FT   CARBOHYD       74     74       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      403    403       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     1023   1023       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     1099   1099       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     1618   1618       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     1737   1737       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     1832   1832       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     1878   1878       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     2136   2136       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     2505   2505       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     2374   2379       NNQKMA -> RAGPGP (IN REF. 1).
FT   NON_TER      2700   2700
SQ   SEQUENCE   2700 AA;  293013 MW;  80E60C0CB12277B1 CRC64;
```

```
Query Match                   7.6%;            Score 555.5;   DB 1;    Length 2700;
Best Local Similarity   32.6%;  Pred. No. 9.1e-15;   DB 1;    Length 2700;
Matches  199;  Conservative  64;  Mismatches 240;  Indels  107;  Gaps   32;

Qy   216 PINPRPSLPPNSDTSKETSLITVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKT 275
             ||||||| |||     ||      |:|        || ::|      ||
Db   445 PVKVLPELPPVSPVSS---TGPSETTGLTENPTISTK----------KPTVSIEKP 487

Qy   276 SAKDLAPTSKVLAKPTPKAETTTKGPALTTPKPEPTPTTPKEPASTTPKEPT----PTTI 330
             || |     ||        |:|  |||         ||:||| |        ||||
Db   488 SVTTEKPT-----VPKEKPTIPTEKPTPISTEK---PTIPSEKPNMPSEKPTIPSEKPTIL 539

Qy   331 KSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTPKEPAPTTTKEPA 389
             ||  ||||||     ||  |||||| ||    ||   ||      ||| :|:|
Db   540 TEKPTIPSE-KPTIPSEKPTISTEKPTIVPTEE--PTTPTEETTYMEEPVIPTEKPSIPT 596

Qy   390 TPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKP 449
             :|||  | ||  |||  ||||| ||        || ||||| ||||||  |||||
Db   597 -EKPSIPTEK---PTISMEETTISTEKPTICPEKPTIPTEK----PTIPTEKSTISPEK 647

Qy   450 APTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKEPAPTTTKEPAPTTKEPAPTTK 509
             |||||||| ||||| ||   |:|  ||        ||:||| |      ||| ||||
Db   648 -PTTPTE-KPTIPTEKPTISTEKPTIPTEK-PTISPEKLITIPTEKLTIPT---EKPTIPT 701

Qy   510 EPSPTTTKEPAPTTTKEPAPTTPKKPA---------PTTPKEPAPTTPKEPAPTTTKKPAPTAP 564
             :|:|        || :|: |            |:|:|      ||| ||||||  |||
Db   702 EKPTISTEE--PTTPTEETTISTEKPSIPMEKPTLPTEETTTISVEETTISTEKLTIPM-- 757
```

RESULT 7

EXTN_TOBAC

ID   EXTN_TOBAC   STANDARD;   PRT;   620 AA.
AC   P13983;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
GN   HRGPNT3.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. XANTHI; TISSUE=Leaf;
RX   MEDLINE=90128263; PubMed=2612909;
RA   Keller B., Lamb C.J.;
RT   "Specific expression of a novel cell wall hydroxyproline-rich
RT   glycoprotein gene in lateral root initiation.";
RL   Genes Dev. 3:1639-1646(1989).
CC   -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC   THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC   MAIN ROOT.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC   -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC   SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC   GLYCOSYLATED.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X13885; CAA32090.1; -.
DR   PIR; S06733; S06733.
KW   Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW   Hydroxylation.
FT   SIGNAL   1   ?
FT   CHAIN   ?   620
FT   REPEAT   70   73   H-A-P-P.
FT   REPEAT   148   151   H-A-P-P.
FT   DOMAIN   229   242   2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT   REPEAT   229   235   1.
FT   REPEAT   236   242   2.
FT   DOMAIN   205   620   CONTAINS THE SER-PRO(4) REPEATS.
FT   DOMAIN   499   600   3 X APPROXIMATE TANDEM REPEATS.

SQ   SEQUENCE   620 AA; 65406 MW; 641DD22278A828524 CRC64;

Query Match   7.6%; Score 551; DB 1; Length 620;
Best Local Similarity   27.2%; Pred. No. 4e-15;
Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

Qy   294 AETTTK-GPALTTP--KEPTTPTT------PKEPASTTPKE---PTPTTIKSAPTTPKEPAP 342
          |:|:|| | |:||| |.||||| ||||| || || || |||| |||||||
Db   24 AEATTQYGGYLPPPVTSQPPPSSIGLSPPSAPTTTPPSRGHVPSP---RHAPPRHAYPPP 80

Qy   343 TTTKSAPTTPKEP------APTTTKEPAPT---TPKEPAPTTTKEPAPTTTKSAPTTPKE 393
          |:| |||||||| ||||| ||| ||||||||||| ||||||||| ||||
Db   81 SHGHLPPSVGGPPPHRGHLPPSRGFENPPPSPVISPSHPPSYGAPPSHGPGCHLPSHGQR 140

Qy   394 PAPTTPKKPAPT-----TPKEPAPTTTPKEPTTTKEPSPTTTK----EPAPTTTKSAPTT 448
          ||||||||||| |||||||||| |||||: :|:|| ::| |||||||||||
Db   141 P----------PSPSHGHAPPSGGGHTPPRGQHPPSHRRPSPPSPPPHGPPPPTTAQP 193

Qy   449 PAPTTTKEPAPTTTKEPAPTTTKEPSPTTTK---EPAPTTTKSAPTTTKEPAPTTTKS-- 503
          ||| ::|:| ||: ||||| |||: |:|:| |||| ||||:| :|||| |||||
Db   194 PSPQV--QPPPT--YSPPPTHVQPPSPSPSPPSRGHQQPPPTHRHAPPTHQPSPL 249

Qy   504 ---APTTPKEPSPTTTKEPAPTTTKEPAPTTTKKPAPTTTPKEPAPTT---TPKEPAPTTT 557
             |||||||||| ||||||||||||||:|||||||| ||||||||| |||||| |||
Db   250 RHLPPSPRRQPQPPPIYSPPPPAYAQSPQSPPTYSPPPPIYSPPPPIYSPPPPAYSPSP 309

Qy   558 KPAPT----APKEPAPTTKETAPTTPKKL-TPTT------PEKLAPTTPEKPAPTTPEEL 607
          |||| |||||||||| :: ||||| ||| |||||||||||||||||||
Db   310 PPTPTPPTFSPPPPAYSPPPPIYSPPPPYLPLPSSPIYSPPPPVYSPPPPVYSPPPPYTL 369

Qy   608 APTTPEEPTPTTPEEPAPTTPKAAAP---NTPKEPAPTTTPKEPAPTTTPKEPAPTTPKETA 664
          |||||:| :|||||||||| |:||| |||:|||||:|||||:|| ||||||||||
Db   370 PPPPPSSPPPPSESPPPPIYEQSPPPPAYSPPLPAPPYSPPPPT--YSPPPPTYAQPP 427

Qy   665 PTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEP 724
          |||:|| ::||| |||||| ||||||||||: ||: :|:||||||||||| |||||
Db   428 P-------LPPTYSPPPPAYSPPPPTYSPPPPISPSRGHQQPPPTHRHAPPTHQPSPPAYSP 481

Qy   725 APTTPKEPAPTTPKGTAPTTPKKPAPKELAPTTKEPAPTTTPKEPAPTT---TPKEPAPTTTK 784
          |:| ||||||||||| | :|||||||| |||||||||||| |||||| ::|||||| |
Db   482 PPPSPIYSPPPPQVQPLPPIFSPPPPRRIHLPPPPHRQP---RPPTPTYGQPPSPPTFSP 538

Qy   785 TAPTTPKEPAPT---TPKKPAPTTPEETPPPPTISEVSTPTTTKEPTTIHKSPDESTPELSAE 842
          :|||||||||| ||||||:|||||||||||| || ||||||||||||:|: |: |||:
Db   539 PPPRQIHSPPPPHWQPRTPTTYGQPPSPPTFSAPPPRQIHSPPPPHRQPRPPTTYGQP 598

Qy   843 PTPKALENSPKEPGVPTIKTPAATKP 868
          |:||:|:|||||| |||:|| |:|||
Db   599 PSP-------PTTYSPSPSPPP 612

```
RL   Gene 68:307-314(1988).
RN   [2]
RX   MEDLINE=89089138; PubMed=3145094;
RP   SEQUENCE FROM N.A.
RA   Shneidman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
RT   "The structure of the largest murine neurofilament protein (NF-H) as
RT   revealed by cDNA and genomic sequences.";
RL   Brain Res. 464:217-231(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER; TISSUE=Brain;
RA   Carden M.J.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.

CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC       NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC       SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC       TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
```

```
DR   EMBL; M24496; AAA39813.1; -.
DR   EMBL; M23349; AAA39813.1; JOINED.
DR   EMBL; M24494; AAA39813.1; JOINED.
DR   EMBL; M24495; AAA39813.1; JOINED.
DR   EMBL; M35131; AAA39809.1; JOINED.
DR   EMBL; Z31012; CAA83229.1; ALT_FRAME.
DR   PIR; JT0368; QFMSH.
DR   PIR; A43778; A43778.
DR   MGD; MGI:97309; Nfh.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Repeat.
FT   DOMAIN        1     97       HEAD.
FT   DOMAIN       98    408       ROD.
FT   DOMAIN      409   1087       TAIL.
FT   DOMAIN      436    517       GLU-RICH (ACIDIC).
FT   DOMAIN      519    886       50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
FT   DOMAIN      887   1087       GLU/LYS-RICH.
FT   DOMAIN       98    129       COIL 1A.
FT   DOMAIN      130    141       LINKER 1.
FT   DOMAIN      142    239       COIL 1B.
FT   DOMAIN      240    261       LINKER 12.
FT   DOMAIN      262    283       COIL 2A.
FT   DOMAIN      284    287       LINKER 2.
FT   DOMAIN      288    408       COIL 2B.
FT   CONFLICT    133    133       K -> QA (IN REF. 2 AND 3).
FT   CONFLICT    199    199       A -> AR (IN REF. 2 AND 3).
FT   CONFLICT    281    281       S -> T (IN REF. 2 AND 3).
FT   CONFLICT    492    492       L -> G (IN REF. 2 AND 3).
FT   CONFLICT    551    551       P -> PREAKSP (IN REF. 3).
FT   CONFLICT    689    712       MISSING (IN REF. 3).
FT   CONFLICT    714    714       G -> A (IN REF. 3).
FT   CONFLICT    814    814       V -> M (IN REF. 2 AND 3).
FT   CONFLICT    843    843       T -> N (IN REF. 2 AND 3).
SQ   SEQUENCE   1087 AA;  116612 MW;  57BAC76A38ED1CB9 CRC64;
```

```
Query Match             7.6%;  Score 551;  DB 1;  Length 1087;
  Best Local Similarity   26.8%;  Pred. No. 6.4e-15;
  Matches  234;  Conservative   96;  Mismatches  358;  Indels  186;  Gaps   43;

Qy    102 KKTKKVIESE--EITEEHSVS-ENQESSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLK 158
              :|:|                     |:||:   |:||:|
Db    331 KSTKESLERQRSELEEDRIRQADIASYQDAIQQILDSELRNTKWEM----AAQLREYQDLLN 385

Qy    159 VK--------DNKKKNTFKKFKPTPKPVVDEAGSGLDNGDFKVTTPDTSTTQHHK 204
             :|                   |:|            ||    |
Db    386 VKMALDIEIAAYRKLLEGEECRICFGFGPSP--------FSLTEGLPKI-PSIST-HIK 432

Qy    205 VSTSPKITTAKPINPRPSLPPNSDTSKETSLITVNKETTVEIFKETTTTNKQTSTDGKEKTT 264
                     |||                                         :|
Db    433 VKSEEMIKVVE----KSEKETVIVEGQTEEIRVTEGVTEEEDKEAQGEGEE 480

Qy    265 SAKETQSIEKTSAKDLAPTSKVLAKPTPKKAETTTKGPALTTPKEPTPTTKEPASTTPKE 324
                                |||                    |||  |
Db    481 AEEGGEEKEEEELAAATSPPAEEAASPEKETKSRVKEEAKSPGEAKSPGEAKSPA----EA 536

Qy    325 PTPTTIKSAPTTPKEPAPTTTKSAPTTPKEP-APTTTKEPAPTTTKKP----- 379
          |  | ||| ||| | :     ||| ||:||   ||| ||| |||
Db    537 KSPGEAKS-PGEAKSPGEAKSPAEPKSPAEAKSPA--EPKSPA--TVKSPGEAK 591

Qy    380 APTTTIKSAPTTPKEPA----PTTPKKPA----PTTTPKEPA----PTTPKEP----TPTTP 423
          : |  | || ||| |        | |         || |         ||| |
Db    592 SPSEAKS-PAEAKSPAEAKSPAEAKSPAEAKSPAVKSPGEAKSPSEA 650

Qy    424 KEPAPTTTKEPAPTTPKEPA----PTAPKKP----APTTPKEPAPTTPKEPAPTTTKEPSP 475
          | |||| |||| ||| |        | |         ||| ||||| ||| ||||  ||||
Db    651 KSPA----EAKSPAEAKSPAEAKSPAEVKSPGEAKSPAEPKSPAEA--KSPA---- 701

Qy    476 TTPKEPA----PTTTKSAPTTTKEPAPTTTKSAPTTPKE-PSPTTTKEPA----PTTPKE 526
           || |        | || |||| ||||| || ||| |  ||| |||| |        ||| |
Db    702 AEAKSPAEVKSPGEAKS-PAAVKSPAEAKSPAAVKSPGEAKSPGEAKSPAEAKSPAEAKS 760

Qy    527 PAPTTTKKPAPTTPKEPAPTTTKEPAPTTKKPAPTAPKEPAPTTPKETAPTTTKKLTPTT 586
          | ||| || ||| | |  || || || ||| | | |  ||| ||:  |||| |  |
Db    761 PIEVKSPEKAKTPVKEGAKSPAEAKSPAEVKSPGEAKSPAEPKSPAEA--KSPVKEGAK 819

Qy    587 TPEKLAPTTPEKPAPTTTPEELAPTTPEEPTPTTPEE-PAPTTPKAAAPNTPKEPAPTT-- 643
           |||  ||| || | ||  |  || ||  | | ||    | || |   | || |
Db    820 PPEKAKPLDVKSPEAQTPVQEEATVPTDIRP--PEQVKSPAKEAKSPE--KEEAKTSEK 875

Qy    644 --PKEPAPTTP------KEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 694
            || | ||        | || || |  | | || | |  ||| || |  | |
Db    876 VAPKKEEVKSPVKEEVKAKEPPKKVEEEKTLPTPKTEAKESKKDEAPKEAPKPKVEEKKE 935

Qy    695 TTTKEPISTTTSD--KPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTT 752
          |  ||     |    | || ||  | || | | || ||| ||| || | | |  ||| |
Db    936 TPTEKPKDSTTAERAKEEAGEKKKAVASEEETPAKLGVKEEA--KPKEKTETTTKTEAEDTK 993

Qy    753 PKKPAPKELAPTTTKGPTSTTSDKPAPTTPKEPAPTTPKKPAPTTPETPPPT 812
          | | |      || |  |  |  | || ||  | ||  |
Db    994 AKEPS----KPTETEKP----------KKEEMPAAPEK---------- 1017

Qy    813 TSEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAATKPEMTT 872
           |  |   ||| | |  ||    | || |  || |     || ||  |||
Db   1018 -----KDTKEEKTTESRKPEEK-----PK-MEAKVKEDDKSLSKEP--SKPKTEK 1059

Qy    873 TAKDKTTERDLRTTPETTTAAPKMTKEFATTTEK 906
          : |  ||| | | || |  | | |    |   ||
Db   1060 AEKSSSTDQKESQPPE-----KTTEDKATKGEK 1087

RESULT  9

TCNA_TRYCR
ID   TCNA_TRYCR     STANDARD;      PRT;  1162 AA.
AC   P23255;
```

```
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SIALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).
GN   TCNA.
OS   Trypanosoma cruzi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5693;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SILVIO X-10/4;
RX   MEDLINE=91277609; PubMed=1711561;
RA   Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
RA   Prioli R.P.;
RT   "The Trypanosoma cruzi neuraminidase contains sequences similar to
RT   bacterial neuraminidases, YWTD repeats of the low density lipoprotein
RT   receptor, and type III modules of fibronectin.";
RL   J. Exp. Med. 174:179-191(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91376547; PubMed=1896773;
RA   Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;
RT   "Trypanosoma cruzi: localization of neuraminidase on the surface of
RT   trypomastigotes.";
RL   Trop. Med. Parasitol. 42:146-150(1991).
CC   -!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
CC       PARASITE INVASION OF CELLS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3- 2,6- AND 2,8-GLYCOSIDIC
CC       LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLNEURAMINYL
CC       RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED
CC       NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
CC       GLYCOLIPIDS OR COLOMINIC ACID.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POSSIBLE).
CC   -!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM
CC       IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
CC   -!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
CC       DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
CC       PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
CC   -!- SIMILARITY: CONTAINS 3 BNR REPEATS.
```

```
DR   EMBL; M61732; AAA30255.1; -.
DR   PIR; JH0557; JH0557.
DR   HSSP; P29768; 1DIL.
DR   InterPro; IPR002860; BNR.
DR   Pfam; PF02012; BNR; 2.
KW   Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
KW   Phosphorylation.
FT   DOMAIN         1    457
FT   REPEAT        23     34       CYS-RICH.
FT   REPEAT       163    174       BNR 1.
FT   REPEAT       209    220       BNR 2.
FT   DOMAIN       458    588       FIBRONECTIN TYPE-III.
FT   DOMAIN       589   1120       44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT   CARBOHYD     342    342       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     394    394       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1125   1125       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1162 AA;  120032 MW;  07049218970C6A40 CRC64;
```

```
Query Match     7.38;    Score 530.5;    DB 1;    Length 1162;
Best Local Similarity   29.4%;    Pred. No. 4.1e-14;
Matches 177;    Conservative 100;    Mismatches 212;    Indels 113;    Gaps 34;
```

```
RESULT 10
CPN_DROME

ID   CPN_DROME     STANDARD;      PRT;   865 AA.
AC   Q02910;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   CALPHOTIN.
GN   CPN OR CAP.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=93165729; PubMed=8094559;
RA   Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT   "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=93165730; PubMed=8434015;
RA   Ballinger D.G., Xue N., Harshman K.D.;
RT   "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT   calcium and contains a leucine zipper.";
```

```
Qy   310 TPTTPKE-PASTTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTP 368
             |:||| :  |  || ::||  |:|| :|| :|||  |:||  :|| ::|| :||
Db   599 TPSTPADSSAHSTPSTPVDS---SAHSTPSTPADSSAHGTPSTPVDSSAHGT----PSTP 651

Qy   369 KE-PAPTTKEPAPTTKSAPTTPKE-PAPTTPKP------APTTPKEPA-----PTTP 415
            :|  :|| :||  || ::|| | :|| :|||       |:||  :||     ||||
Db   652 ADSSAHGTPSTPVDSSAHSTPSTPADSSAHSTPSTPVDSSAHSTPSTPADSSAHGTPSTP 711

Qy   416 KEPTP-TTPKEPAPTTKEPAPTTKEPAPTTTKKPAPTAPKEPAPTTPKEPAPTTKEPAPTTPKS 474
            ::|| :|| :||  || :|||  || ::|| :|| ::||  :|| ||| ::||  || 
Db   712 VDSSAHGTPSTPVDSSAHGTPSTPVDSSAHSTPSTPADSSAHSTPSTPADSSAHSTPSTP 771

Qy   475 PTTPKEPAPTTKSAPTTTKEPAPTTTKEPAPTTPKEPSPTTKEPAPTTPKEPA----PTT 530
            :|| :||  ||  :||  :|| :||  :|| ||| :|| :|| :||  || ||    || 
Db   765 HGTPSTPADSSAHSTPST----PADSSAHGTPSTPVDSSAHSTP----PSTPVDSSAHGTPS 817

Qy   531 TP-KKPAPTTPKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKL----TP 585
            || :|| ::|| :|| ::|| :||  || ::|| :|| ::|| ::||  || ||||  ||
Db   818 TPVDSSAHSTPSTPVDSSAHGTPSTPVDSSAHSTPSTPADSSAHST-PSTPADSSAHGTP 876

Qy   586 TTP-EKLAPTTPEKPAPTTPEELAPTTPEE----PTPTTPEEPA----PTTP-KAAAPNT 635
            :|| :||  :|| :||  ::|| :||  ||    ||| :|| ||    |||| :||  ||
Db   877 STPVDSSAHSTPSTPADSSAHS-TPSTPVDSSAHGTPSTPADSSAHGTPSTPVDSSAHGT 935

Qy   636 PKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKE-PAPTTPKEPAPTTPKPAPKELAPT 695
            | :||  || :||  || :|| :|| ::|| | :|  :||  ::||  || ||| :|||
Db   936 PSTPADSS-----AHSTPSTPADSSAHST-PSTPADSS------AHSTPSTPVDSSAHST 983

Qy   696 TTKEPTSTTSDKPAPTTPKGTAPTTPKE-PAPTTPKEPAPTTPKGTAPTTLKEPAPTTPK 754
            :|| :|| ||  |:|| || ::|| | :| |:|| :|| :|| || ::|| |:|| :|||
Db   984 -----PSTPADSSAHST----PSTPADSSAHSTP----AHSTPSTPVDSSAHSTPS 1033

Qy   755 KPAPKELAPTTTKGPTSTTSDKPAPTTPK----ETAPTTKEPAPTTPKKPAPTTPETPP 810
            :|| ::|| :|| | :|| ||  |:|| ||    ::|| :|| :|| ::|| :|| :||
Db  1034 TP------ADSSAHGTPSTPADSSAHSTPSTPVDSSAHSTPSTPADSS----AHGTPSTPA 1084

Qy   811 PTTSEVSTPTTTKEPTIIHKSPDESTP-ELSAEPTPKALENSPKEPGVPTTKTPAATKPE 869
            ||| : || :|| :| :|:|  |:||  :|| :|| | ::|| :|| ||| :||| |:|
Db  1085 DSSAH-STPSTPAD-SSAHGTP--STPADSSAHSTP----------------STPAGSSAN 1125

Qy   870 MT 871
            |:
Db  1126 GT 1127
```

RL   Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
-!- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC   REGULATE THE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC   OF CA+2 PER MOL OF PROTEIN.
CC   -!- SUBUNIT: HOMODIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC   -!- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC   COMPOUND EYES AND OCELLI.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC   DEVELOPMENT.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; L02111; AAA28405.1; -.
DR   EMBL; L05080; AAA28420.1; -.
DR   PIR; A47282; A47282.
DR   FlyBase; FBgn0010218; Cpn.
KW   Calcium-binding.
FT   CONFLICT   36   36   A -> AVAPAVVA (IN REF. 2).
FT   CONFLICT   43   43   I -> T (IN REF. 2).
FT   CONFLICT   64   64   I -> V (IN REF. 2).
FT   CONFLICT   76   76   I -> A (IN REF. 2).
FT   CONFLICT   100   100   P -> PP (IN REF. 2).
FT   CONFLICT   126   127   VQ -> AP (IN REF. 2).
FT   CONFLICT   154   154   I -> V (IN REF. 2).
FT   CONFLICT   160   160   S -> T (IN REF. 2).
FT   CONFLICT   534   534   A -> E (IN REF. 2).
FT   CONFLICT   699   699   I -> T (IN REF. 2).
FT   CONFLICT   703   703   V -> L (IN REF. 2).
FT   CONFLICT   721   721   D -> E (IN REF. 2).
SQ   SEQUENCE   865 AA; 84781 MW; 21104I7E0B0E7CFE CRC64;

Query Match   7.3%; Score 530; DB 1; Length 865;
Best Local Similarity 26.2%; Pred. No. 3.4e-14;
Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;

Qy   281 APTSKVLAKP-TPK---AETTTKRGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTT 336
          : | :|| | || | ||| | |||| ||| :|| ||| |||| |||||| |||
Db   8 SPVSAPVAAPVTPSAVAAPVQVSPAAVAP-------APAAPIAVTVAPPPTLASVQPAT 61

Qy   337 PKEEAPTTTKSAPTTP-KEPAPTTTKEPAPTTPKEPAPTTTKEPAPTT----TTTKSAPTTP 391
          ||| | | |||| ||| ||| || ||| || ||| ||||||||||
Db   62 VTIPAPIAAASVTPVASAVAPPVVAAPTP----PRAASPVSTPVAVAQIPVAVSAPVAP 116

Qy   392 KEPAPTTPKKPAP--------TTTKEPAPTTPKEPTP-----------TTTPKEPAPTTKEPAPTT 437
          | | ||| || | || |||| | ||| | || |||| |||| |||
Db   117 PVAAPTPVVQIPVAAPVIATPPVAASAPTPAAVTPVISPVIASPPVVPANTIVPVAAPV 176

Qy   438 PKEPA--PTAPKKPAPTTPKEPAPTTPKEPAPTTPKE-PSPTTPKEPAP---- TTTKSA 489
          ||||| || || ||||||| ||| ||| ||| | ||||||| |||| |||||
Db   177 AAVPAAVPVVAPVLAPAVAPAVAVVAETPAPPPVVAEIPVATIPECVAPLIPEVSVVATK 236

Qy   177 AAVPAAVPVVAPVLAPAVAPAVAVVAETPAPPPVVAEIPVATIPECVAPLIPEVSVVATK 236

Db   237 PLAAAEPVVVAPPAPAETPVVAPAAAASPHVSVAPAVETAVVAPVSASTEPPVAAATLTTAP 296

Qy   490 PTTTKEP--APTTTKSAPTTPKEPSPTTTKEPA-PTTPKEPAPTTPKKPAPTTPKEPAP 545
          || ||| |||| || |||| || |||| | |||||||| || | |||| || ||

Db   546 TTP-----------KEPAPTTTKKPAPTTPAPKEP------ APTTPKET------------- APTTPKKLTPTT 586

Qy   297 ETPALAPVVAESQVAANTVVATPTPAPEPEETTAPPVVAETPEVASAVSAVAETTPEVVPPV 356

Db   587 TPEKL-APTTPEKPAPTTPEELAPTTPEEPTTTPEEP---APTTPKAAAPNTPKEPAPTT 642

Qy   587 TPEKL-APTTPEKPAPTTPEELAPTTPEEPTTTPEEP---APTTPKAAAPNTPKEPAPTT 642

Db   357 AAESIPAPVVAATTPVPAT---LAVTDPDVIASAVPELPVVIAPSPVPSAVAETPVDLAPP 413

Qy   643 TPKEPAPTTPKEPAPTTPKEPAPTTP-KGTPAPTTLKE--------------PAPTTPKPAPKE 691
          ||| | | ||| ||| ||

Db   414 V-----LPPVAAEPVPAVVAEETPETPAPASAPVTIAALDIPEVAPVIAAPSDAPAEAPSA 469

Qy   692 LAPTTTKEPTSTTTSDKPAPTTPKGTPAPTTP------KEPAPTTPKEP-APTTPKGTAPT 743
          ||| |||| |||||| | ||||| | | |||| ||||||

Db   470 AAPIVSTPP--TTASVPETTAPPAAVPTEPIDVSVLSEAAIETPAVPPVEVTTEVAVADV 527

Qy   744 TLKEPA-----------PTTPKKPAPKELAPTTP----KGPTSTTSDKPAPTT---PKETAPTT 789
          ||||| || ||| || ||| |||| |||| | ||| ||||

Db   528 APPEAAADLLIEPVEPPAPIPDLLEQTTSVPAVEAAESTSS--PIPETSLPPPNEAVASP 585

Qy   790 PKEPAPTTPKKPAPTTPETTPPTTTSEVSTPTTTTKEPTTTIHKSPD------------------ESTPEL 839
          |||| ||| | || || || | |||| ||||| ||| ||| |||||

Db   586 EVAVAPITAPEIP----EPPPSLATPTEPIPVEAPVVIQEAVDAVEVPVTETSTSIPET 641

Qy   840 SAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTP--ETTTAAPKMT 897
          | ||| | | |||| | || ||| ||||||| |||||||||

Db   642 TVE-FPEAVAEKVLDPAI--TEAPVTTQEPDVANINDGAPATEI-TTPAVELVTAAAEVS 697

Qy   898 KEPATTTEKTTTESKITATTTQVTST------TTQDTTPFKITTLKTTTLA-PKVTTTKKT 950
          ||| | | | || ||| ||||| || ||| | ||||||

Db   698 DIAIPVIDPPVPQEIAVAEIPETDTKPAEVIVEQSTIPIEAPVPEVSKYAEPVVISEAPAA 757

Qy   951 ITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRKPKTTPT 1010
          | | || | | | | | | | |||||||||

Db   758 EVPITAGDNPDNTSVGLSEVVPTIAEKPVEEVPTS--EIPEQSSSPS--DSVPVAKITPL 813

Qy   1011 PRKMTSTMPELNPTPSRIAEAM---LQTTTRPNOTPNSKLVEVN-------PKS 1053
          | || | || | | | ||| | | |||

Db   814 LRDLQTTTDVSLLAIAATLDAIGEKLKDQKARNQQVMDRLCEIEKILGPPKKS 864

RESULT 11
ID   FP1_MYTCO   STANDARD;   PRT;   872 AA.
AC   Q25434;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCFP1).
GN   FP1.
OS   Mytilus coruscus (Sea mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=42192;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foot;
RX   MEDLINE=96394686; PubMed=8798340;
RA   Inoue K., Takeuchi Y., Takeyama S., Yamaha E., Yamazaki F., Odo S.,
RA   Harayama S.;
RT   "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and
RT   its evolutionary implications.";
RL   J. Mol. Evol. 43:348-356(1996).
-!- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC   PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC   ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC   FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC   -!- DOMAIN: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND
CC   ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).
CC   -!- PTM: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; D63777; BAA09850.1; -.
DR   InterPro; IPR002964; Adhesive_plaq.

```
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01216; ADHESIVEI.
     PRINTS; PR01217; PRICHEXTENSN.
KW   Signal; Repeat; Hydroxylation.
FT   SIGNAL         1     20       POTENTIAL.
FT   CHAIN         21    872       ADHESIVE PLAQUE MATRIX PROTEIN.
FT   DOMAIN        21     41       NONREPETIVE LINKER.
FT   DOMAIN       124    872       TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-
FT                                 P-[PST]-[ST].
FT   DOMAIN       184    192       NONPEPTIDE 1.
FT   DOMAIN       213    221       NONPEPTIDE 2.
SQ   SEQUENCE   872 AA;  101677 MW;  98CC70D7C75FF3C4 CRC64;
```

```
Query Match            7.1%;
Best Local Similarity  29.2%;             Pred. No. 9.3e-14;
Matches  276;  Conservative  80;  Mismatches 407;  Indels 181;  Gaps 54;
                                          Score 518.5;  DB 1;  Length 872;
```

```
Qy    177  VVDEAGSGLDNGDFKVTTPDTST---TQH------NKVSTSPKITTAKPINPRPSLPPNS  227
                :|  |                         ::|
Db     29  VYGSAVSGASAGAYK-TLPGSHPYGSKHVPVYKPMNKIPT-PYI--SKKSYPAYKPKGY   84

Db     85  YPTYKRYQPTYGSKTNPYIYKPIAKLSSYKAIKTTYPAYKAKTSYPPSYK-----HKIT  139
                  ||  :|
Qy    228  DTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVL  287

Qy    288  AKPTPKAETTTKGPALTTPKEPT-PTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTYK  346
                  :|| |          ::| |:  |          ||
Db    140  YPPTYK------PKITYP--PTYKQKPSYPPSYKPKITYPP-----PKITYPPTYK     185

Qy    347  SAPT-TPKEPAPT--TTKEPAPTTPKEPAPTTTPKEPAPTTYK--PKITYPPTYK--PKK  401
                |  |       ||          ||  |          ||
Db    186  RKPSYPPSYPYKPKATYPPTKITYP-----PTYKRPSYT-----PYKPKITYPPTYKPKPKI  236

Qy    402  PAPTTPKEPAPTT----PKEPTPTT--KEPAPTTPKEPAPTTTPKEPAPTTP-APKKPAPTTP  454
                     |:|               :|             ::|
Db    237  SYPSIYKPKASYVSSVSKKKTYPPIYKPKISYPPTYKPKPSYP----PTYKPKTYPPPTY    292

Qy    455  KEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTKSAPT-------TTKEPAPTTKSAPTTPK  509
                :|          |            :|              ||
Db    293  K-PKPSYPPTYKPKITYPPT-YKPKSYPPIYKSKSSYPTSYKSKKTYP             350

Qy    510  EPSPTTTKEPA--PTTPKEPAPTTPTPKKP------EPAPTTPKEPAPTTTKKP       559
                |            :|          |
Db    351  TYKPKITYPPTYKPKPSYPPSYPPSYPKKTYSPTTYKPKIYPPTYKPKSYPPSYKPKITYPP  410

Qy    560  APTAPKEPAPTTPKETAP------TTPKKLTTPTTPEKLAPTTPEKPAPTTPEELAPTTPEE  614
                  |:| |                  |           ||
Db    411  T-YKPKISYPPTYKPKASYVSVSKKKTYPPIYKSYPPTYKPKPSYPPTYKPKITYP      469

Qy    615  PT-PTTPEEPAPTTPKAAAPNTPK-EPAPTTPKERAPTTP-----KEPAPTT--PKETAP  665
                   |:|         |           :|              ||
Db    470  PTYKPKPSYPPTYKPKITYPPTYKPKISYPPTYKPKITYPPTYKPKQKPSYPPTY       529

Qy    666  TT--PKGTAPTTLKEPAPTTPKKPAPKELAPTTTKPKEPTSTISDKPAPT-TPKGTAPTTPK  722
                :|           ||             |:|
Db    530  PTYKPKITYPPTYK-----PKPSYPPSYKPKITYPPTYKPKIRYPPTYKPKRASYPPTYK   583

Qy    723  EPAPTTKEPAPTTPKGTAPTTLKEPAPTTP------KKPAPKELAPTTTKGPTSTSDKPA  778
                 |:|                  ||            |:|
Db    584  PKITYPPTYK-PKPSYPPTYKQKPSYPPTYKSKSSYPTAYKSKKTY               628

Qy    779  PTT--PKETAPTTPKEPAPTPK--KPAPTTPET--------PPPTSEVSTPPTTTKEPTT  827
                  |            |:|              ||
Db    629  PPTYKPRKITYPPTYK-PKPSYPPSYRKIIYPPTYKKSYPQAYKSKGSYPPSYQPKKT    687

Qy    828  IHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAAATKPEMTTTAKDKT----TERDL  883
                :|                   |:|
Db    688  Y-----PPSYKPKITYPPTYK----TKPSYPASYKRKTYSYPPTYKPK            735

Qy    884  RTTPEPTTTAAPKMTKETATTTEKTTESKIPATTTQVTSTTTQDDTTPFKITTLKTTTLAPK  943
                :|               ||            |:|
Db    736  ISYPSTYKAKP----SYPPTYKPKPSYASSYKPKIRYPPTYKPKPSYASSYK----PK   785
```

```
Qy    944  VTTTTKKITTTEIMNKPEETAKPKDRATNSKATTPKPQ---KPTKAPKKP-TSTKPK-T  998
               :|| |                          ||
Db    786  I---------------------RYPPTYKPKP----SYASSYKPKIRYPPTTYKPKPSYASSYKPKIT  827

Qy    999  MPRVRKPKTT--PTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQ 1040
               :|                                 ||
Db    828  YPPTYKPKISYPPTTYKPKIRYPPTTYKPKISYPPPAYKPKISYPSQ 871
```

```
RESULT  12

ID   NFH_RAT      STANDARD;      PRT;   831 AA.
AC   P16884; Q63368;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE   (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
GN   NFH OR NFH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=89065087; PubMed=3143606;
RA   Breen K.C., Robinson P.A., Wion D., Anderton B.H.;
RT   "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT   Identification of putative phosphorylation sites.";
RL   FEBS Lett. 241:213-218(1988).
RN   [2]
RP   SEQUENCE OF 37-831 FROM N.A.
RX   MEDLINE=88309090; PubMed=2457365;
RA   Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
RA   Jolles P.;
RT   "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT   in situ detection.";
RL   Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RN   [3]
RP   SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX   MEDLINE=87080760; PubMed=2878828;
RA   Robinson P.A., Wion D., Anderton B.H.;
RT   "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT   (NF-H).";
RL   FEBS Lett. 209:203-205(1986).
RN   [4]
RP   SEQUENCE OF 318-831 FROM N.A.
RX   MEDLINE=89184647; PubMed=2928342;
RA   Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
RA   Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
RT   "Cloning of a cDNA encoding the rat high molecular weight
RT   neurofilament peptide (NF-H): developmental and tissue expression in
RT   the rat, and mapping of its human homologue to chromosomes 1 and
RT   22.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC       NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC       SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE INTERMEDIATE FILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC       ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
```

```
DR   EMBL; M37227; AAA41693.1; ALT_FRAME.
DR   EMBL; X13804; CAA32038.1; ALT_FRAME.
DR   EMBL; M21964; AAA41695.1; -.
DR   EMBL; J04517; AAA41692.1; -.
DR   PIR; A30796; A30796.
DR   PIR; A25649; A25649.
DR   PIR; B25649; B25649.
DR   PIR; S02003; S02003.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
FT   NON_TER         1      1
FT   DOMAIN        276    641       51 X 3 AA TANDEM REPEATS OF K-S-P.
FT   CONFLICT      164    164       L -> I (IN REF. 2).
FT   CONFLICT      185    185       L -> S (IN REF. 2).
FT   CONFLICT      193    193       L -> T (IN REF. 2).
FT   CONFLICT      199    199       M -> T (IN REF. 2).
FT   CONFLICT      346    346       K -> N (IN REF. 1).
FT   CONFLICT      373    373       A -> V (IN REF. 1 AND 4).
FT   CONFLICT      482    482       G -> E (IN REF. 2 AND 4).
FT   CONFLICT      485    485       P -> S (IN REF. 2).
FT   CONFLICT      570    570       RK -> KE (IN REF. 2 AND 4).
FT   CONFLICT      591    591       P -> T (IN REF. 2 AND 4).
FT   CONFLICT      727    727       A -> V (IN REF. 4).
FT   CONFLICT      757    759       AAP -> GST (IN REF. 4).
FT   CONFLICT      769    769       T -> L (IN REF. 2).
FT   CONFLICT      775    775       R -> P (IN REF. 2 AND 4).
SQ   SEQUENCE      831 AA;  89486 MW;  1B0973C3F13EF768 CRC64;
```

```
Query Match             7.1%;
Best Local Similarity   26.6%;     Score 517.5; DB 1; Length 831;
Matches   217;  Conservative    76;  Pred. No. 9.8e-14;
Mismatches  345;  Indels  179;  Gaps   41;
```

```
Qy   102  KKTKKVIESE--EITEHSVS-ENQESSSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLK  158
          |  ||| ||          |:  |   : :||||||||||||| |: || :  |  :|| |
Db   103  KSTKESLERQSELEDRHQVDMASYQDAIQQLDNELRNTKWEM----AAQLREYQDLLN     157

Qy   159  VK-------------DNKKNETKKKPTKPKPPVVDEAGSGLDNGDKVTTPDTSTTQHNK     204
          ||             :  |  |  || |  |    :|       | | |  |  :| |
Db   158  VKMALDLEIAAYRKLLEGEECRIGFGPIP-----------FSLTEGLPKI--PSMST--HIK  204

Qy   205  VSTSPKITTAKPIPNPRPSLPPNSDTSKETSLTIVNKETTVEIRKETTTINKQTSTDGKEKTT  264
          : |||                  |  :|||   :   :| :||:              ||
Db   205  -------------KSEKETIVIEEQTEEIQVTEEVTEEEDKEAQGEEEEE               252

Qy   265  SAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKEP----TPTTPKEPA-S    319
          : :|  |   | |    |  |    |||  :   |   :||||     | || | |
Db   253  AEEGGEEAATTS----PPAEEAASP----EKETKSPVKEEAKSPAEAKSPAEA          303

Qy   320  TTPKEPTPTTIKSAPTTPKEPA----PTTPKSAPTTPKEPA----PTTTKEP---APTT    367
          |||   | || |  |||| |      || | | ||| |      ||| ||     ||
Db   304  KSPAEVKSPAVAKSPAEVKSPAEAKS-PAEAKSPAEVKSPGEAKSPAE                362

Qy   368  PKEPA-----PTTTKEPA----PTTTKSAPTTPKEPA----PTTPKKPA----PTTPKEPA  411
          || |      ||| | |     ||| | | ||| |      ||| | |     ||| | |
Db   363  AKSPAEVKSPAEAKSPAEAKSPASVKS-PGEAKSPAEAKSPAEVKSPATVKSPVEAKSPA   421

Qy   412  ----PTTPKEP---TPTTPKEPAPTTPKEP----APTTPKEPA----PTTAPKPAPTTPK    455
              ||| | |    | ||| | ||| | |     ||| | |      || || | |||
Db   422  EVKSPVTVKSPAEAKSPVEVKSPA-SVKSPEAKSPAGAKSPAEAKSPVVAKSPAEA-K     478

Qy   456  EPAPTTPKEPAPTTPKEPSPTTPKEPA----PTTTKS-----APTTTKEPAPTTTKS---    503
          | ||| | | ||| | | | ||| |      ||| |       ||| | | ||| |
Db   479  SPAGAKPPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPVEVKSPEAKSPVKEGAKSLAE    538

Qy   504  --APTTPKEPSPTTKEPA-----PTTPKEPAPTTPKKPAPTTPKEPAPTTTK         557
            ||| | | | ||| |       ||| | | ||| | | ||| | | ||||
Db   539  AKSPEAKSPVKEEIKPPAEVKSPMRKEAKSPEKAKTLDVKSPEAKPPAEEAK          598

Qy   558  KPAPTAPKEPAPTTPKETTAPTTPKKLITPTTPEKLAPTTPEKLAPTTPEELAPTTPEEPTP  617
          :| |  | ||   | |    ||  | |  ||:|   | ||  | | |       :|
Db   599  RPADIRSPEQVKSPAKEEAKSPEKEETRT--EKVAPKKEEVKSPV--EEV---KAKEPPK    651

Qy   618  TTPEEPAPTTPKAAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLK  677
          |||  | || | |   ||| | | ||| | | ||| | | ||| | | ||| | | || |
Db   652  KVEEEKTPATPKTEVKESKKDEAPKEAQKP-KAEEKEPLIEKPKDSPGEAKKEEA----K  706

Qy   678  EPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTP  737
          | ||| | | ||  | ||| | | ||| | | ||| | | ||| | | ||| | | |||
Db   707  EKKAAAPEEETPAKLGVKEEAKPKEKAEDAKA-----------KEPSKPSEKE-----KP  750

Qy   738  KGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTT  797
          |   | || | ||| | | ||  | ||| |       | | ||| | | ||| | | ||
Db   751  K------KEEVPAAPEKKDTKE-------------EKTTESSKKREEKPKMEAKAKEED----  789

Qy   798  PKKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDE  834
          | | | ||    ||:|  | | |||:| |:|||| |
Db   790  --KGLPQEPSKPKTEKAEKSSSTDQKDSQPSEKAPED  824
```

```
RESULT 13
MUC1_XENLA
ID   MUC1_XENLA     STANDARD;      PRT;   662 AA.
AC   Q05049;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   INTEGUMENTARY MUCIN C.1 (FIM-C.1) (FRAGMENT).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=93077556; PubMed=1447205;
RA   Hauser F., Hoffmann W.;
RT   "P-domains as shuffled cysteine-rich modules in integumentary mucin
RT   C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic
RT   polymorphism.";
RL   J. Biol. Chem. 267:24620-24624(1992).
CC   -!- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC       INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC       MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SKIN.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   -!- SIMILARITY: CONTAINS 6 P-TYPE (TREFOIL) DOMAINS.
```

```
DR   EMBL; L02115; AAA74725.1; -.
DR   PIR; A45155; A45155.
DR   HSSP; P04002; 1WFA.
DR   InterPro; IPR000519; P_trefoil.
DR   Pfam; PF00088; trefoil; 6.
DR   SMART; SM00018; P; 6.
```

```
DR   PROSITE; PS00025; P_TREFOIL; 6.
KW   Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN        1    144      ALA/THR-RICH.
FT   REPEAT       81     88      8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT   REPEAT       89     96      1-1.
FT   REPEAT       97    104      1-2.
FT   REPEAT      105    112      1-3.
FT   REPEAT      113    120      1-4.
FT   REPEAT      121    128      1-5.
FT   REPEAT      129    136      1-6.
FT   REPEAT      137    144      1-7.
FT   DOMAIN      161    202      1-8.
FT   DOMAIN      218    301      P-TYPE 1.
FT   DOMAIN      218    301      8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT   REPEAT      218    224      2-1.
FT   REPEAT      225    239      2-2.
FT   REPEAT      240    249      2-3.
FT   REPEAT      250    259      2-4.
FT   REPEAT      260    275      2-5.
FT   REPEAT      276    287      2-6.
FT   REPEAT      288    294      2-7.
FT   REPEAT      295    301      2-8.
FT   DOMAIN      306    347      P-TYPE 2.
FT   DOMAIN      353    394      P-TYPE 3.
FT   DOMAIN      402    522      12 X APPROXIMATE TANDEM REPEATS,
FT                               THR-RICH.
FT   REPEAT      402    411      3-1.
FT   REPEAT      412    419      3-2.
FT   REPEAT      420    431      3-3.
FT   REPEAT      432    443      3-4.
FT   REPEAT      444    453      3-5.
FT   REPEAT      454    460      3-6.
FT   REPEAT      461    472      3-7.
FT   REPEAT      473    479      3-8.
FT   REPEAT      480    491      3-9.
FT   REPEAT      492    498      3-10.
FT   REPEAT      499    515      3-11.
FT   REPEAT      516    522      3-12.
FT   DOMAIN      516    566      P-TYPE 4.
FT   DOMAIN      525    613      P-TYPE 5.
FT   DOMAIN      572    661      P-TYPE 6.
FT   DISULFID    162    188      BY SIMILARITY.
FT   DISULFID    172    187      BY SIMILARITY.
FT   DISULFID    182    199      BY SIMILARITY.
FT   DISULFID    307    333      BY SIMILARITY.
FT   DISULFID    317    332      BY SIMILARITY.
FT   DISULFID    327    344      BY SIMILARITY.
FT   DISULFID    354    380      BY SIMILARITY.
FT   DISULFID    364    379      BY SIMILARITY.
FT   DISULFID    374    391      BY SIMILARITY.
FT   DISULFID    526    552      BY SIMILARITY.
FT   DISULFID    536    551      BY SIMILARITY.
FT   DISULFID    546    563      BY SIMILARITY.
FT   DISULFID    573    599      BY SIMILARITY.
FT   DISULFID    583    598      BY SIMILARITY.
FT   DISULFID    593    610      BY SIMILARITY.
FT   DISULFID    621    647      BY SIMILARITY.
FT   DISULFID    631    646      BY SIMILARITY.
FT   DISULFID    641    658      BY SIMILARITY.
FT   VARIANT     276    276      K -> E.
FT   VARIANT     354    354      C -> R.
FT   VARIANT     415    415      T -> A.
SQ   SEQUENCE   662 AA;   67774 MW;   F08527F1ED2FD40 CRC64;
```

```
Query Match       7.0%;
Best Local Similarity   28.1%;
Matches  223;    Conservative  48;
Score 509.5;  DB 1;  Length 662;
Pred. No. 1.6e-13;
Mismatches  276;    Indels  247;    Gaps  30;

Qy   250 TTINKQTSTDGGEKTTSAKETQSTEKTSAKDLAPTSKVLAKPTPKAETTKGPALFTPKEP 309
         |||  |  ::||  ::||::|  |||  | |||:  ::|:|| |||||||:|
Db     3 TTAAAVAATGKDTTAAAEGSAAAEKTAA-----AGEVSAPPT--AAVAATGEDATT----- 51
```

```
RESULT 14
ID   RPB1_HUMAN     STANDARD;     PRT;   1970 AA.
AC   P24928;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).
GN   POLR2A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92178992; PubMed=1542581;
RA   Wintzerith M., Acker J., Vicaire S., Vigneron M., Kedinger C.;
RT   "Complete sequence of the human RNA polymerase II largest subunit.";
RL   Nucleic Acids Res. 20:910-910(1992).
```

```
Qy   310 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTPK 369
         ||||     |||        ||||       ||||         |||
Db    52 ------AAAATAAAEETTAAAGEAPTTTAPATTAAGKAPTTAAATAAPTTAT 103

Qy   370 EPAPTTTPKEPAPTTTKS-APTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPTPTTPKEPA- 427
         ||||||        ||||     ||||         ||||          ::
Db   104 GKAPATAAAPVPTTAASKAPTTAAAATHSTAAANAPTTAASAAAKSKERSRTSSSSEEEHCH 163

Qy   428 --PTTKEPAPT---TPKEPAPTTAPKKPAPTTPK------------EPAPTTPKEPAPTTKE 472
           |: |       ||   ||  |        |:             ||||||||||||
Db   164 VKPSKREMCGSKGITKKQ------CKKKNCCPDPEKGHGGIHCFHRKPKGHSHEEHTTTK- 218

Qy   473 PSPTTTPKEPAPTTTKSAPTTTKEPAPTTTKEPSPTTPKEPAPTTTPKEPAPTTP 532
         |||         |||            |||                   ||||
Db   219 ---APTTIQIATTTTT---TPTTT-------------TTTTKATPTT--------- 245

Qy   533 KKPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLA 592
         ||           ||||             |||
Db   246 ----TTTTKATPTTT---------TTTKATPTTT-------------- 270

Qy   593 PTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAAPNTPKEPAPTTPKEPAPTTP 652
         |||                                              ||||
Db   271 ----TTTTKATTTTPTTTTTTTTP----------------- 289

Qy   653 KEPAPTTPKETAPTTPKGTAPTTLKEPAPTTKKPAPKELAPTTTKEPTSTTSDKPAPTT 712
         |||            |||                |||
Db   290 --------TTTTKATTTTTTTSGEECKMEPSK---------REDCGYSGTTESQCR 328

Qy   713 PKG-----TAPTT-----PKEPAPTTPKEPTTPKGTAPTTLKEPAPTTPKPAPKELA 762
         |        |          ||||                   ||||
Db   329 TKGCCPFDSSIPQTKWCFYTLSQVADCKVEPSQRVDCGFRGIT-----ADQCRQKNCCPDSS 384

Qy   763 PTTTKGPTSTTSDKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPETPPPTTSEVSTPTTT 822
         |||                     ||||               ||||
Db   385 ISGTKWCFYSTSQVAA--TKTTTTPTTTTTTTTTTTTTTTTTKATTTTP----- 436

Qy   823 KEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTTKTPAATKPEMTTTTAKDKTTERD 882
         |||                        |||
Db   437 TTTTT----------------TTKA------------------KAT--- 463

Qy   883 LRTTPETTTPAAPKMTKETATTTEKTTESKITPATTTQVTSTTTQDTTPFKITTFLKTTTLAP 942
         |||                  |||:                  |||
Db   464 -TTTPTTTTTP-----------TTTTTKAT-------------------- 510

Qy   943 KVTTTTKKITTTTEIMNKPEETAKPKDRATNSKATTTPKPQKPTKAPKKPTSTKKPK---T 998
         |||                        ||||                :: |||
Db   511 TTTTTKATTTTTSGECKME-------PSKRADCGYPGITESQCRSKGCCPFDSSIPQTKWCFYS 566

Qy   999 MPRVRKPKTTPTPPR 1012
         ||:|    |:|
Db   567 LPQVADCKVAPSSR 580
```

RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95347616; PubMed=7622068;
RA   Mita K., Tsuji H., Morinyo M., Takahashi E., Nenoi M.,
RA   Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
RT   "The human gene encoding the largest subunit of RNA polymerase II.";
RL   Gene 159:285-286(1995).
-!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC   OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC   SUBSTRATES.
-!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC   RNA(N).
-!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
-!- SUBCELLULAR LOCATION: NUCLEAR.
-!- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC   THE PHOSPHORYLATION ACTIVATES POL2.
-!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC   FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC   PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC   III FOR 5S AND TRNA GENES.
-!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

DR   EMBL; X63564; CAA45125.1; -.
DR   EMBL; X74874; CAA52862.1; -.
DR   EMBL; X74873; CAA52862.1; JOINED.
DR   EMBL; X74872; CAA52862.1; JOINED.
DR   EMBL; X74872; CAA52862.1; JOINED.
DR   EMBL; X74871; CAA52862.1; JOINED.
DR   EMBL; X74870; CAA52862.1; JOINED.
DR   PIR; S21054; S21054.
DR   MIM; 180660;
DR   InterPro; IPR000684; RNA_polII_repeat.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR002879; RNA_pol_A2.
DR   Pfam; PF00623; RNA_pol_A; 1.
DR   Pfam; PF01854; RNA_pol_A2; 1.
DR   Pfam; PF00103; RNA_POL_II_REPEAT; 43.
DR   PROSITE; PS00115; RNA_POL_II_REPEAT; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW   DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT   ZN_FING   71   87   C2H2-TYPE (POTENTIAL).
FT   DOMAIN   1590   1958   CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT   CONFLICT   1067   1067   W -> L (IN REF. 2).
FT   CONFLICT   1449   1449   D -> Y (IN REF. 2).
SQ   SEQUENCE   1970 AA;   217205 MW;   6876FC25692A657E CRC64;

Query Match   6.9%;
Best Local Similarity   33.3%;   Score 503.5; DB 1;   Length 1970;
Matches 181; Conservative 101; Pred. No. 6.9e-13;
Mismatches 167; Indels 95; Gaps 47;

Qy   309 PTPTTPKEPASTTPKE-PTPTTIKSAPTTPKEPAPTTTKSAPTTPKE-----PAPTTKE 362
          :|| ||||| || ||| ||| |||:| ||||
Db   1507 PSPMGGISPAMTPWNQGATPAYGAWSPSVGSGMTPGAAGFSPSAASDASGFSGYSPAWS 1566

Qy   363 PAPTTPKEPAPTTPKKEPAPTTTKPEPAPTTPKEP--APTTPKEPTP 420
          |:| |:|: || ||| |||||||:|
Db   1567 PTPGSPGSPGGAMSPSYSPTSPA-YEPRSPGGYTPQSPSPTSPSYSP-SP 1624

Qy   421 TTPKEPAPTTPKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTTTKEPSPTTPKE 480
          :|||| |:| || || |||||||:||| |||:|
Db   1625 TSPSY-SPTSPNYSPSPSY-SPTSPSY-SPTSPSY----SPTSPSY 1671

Qy   481 PAPTTTKSAPTTTKEPAPTTTKEPAPTTPKEPSPTTPKEPAPTTPKEPAPTTPKEPSPTSP 540
          :||: ||:| ||:| ||| |||| |:|:|
Db   1672 -SPTSPSYSPTSPSY-SPTSPSY-SPTSPSY--SPTSPSY-SPTSP 1725

Qy   541 KEPAPTTPKEPAPTTPKEPAPTTTKKPAPTAPKKEPAPTTTPKETAPTTPKKLTPTTPEKLAPTTPEKPA 600
          :||:| ||:| :||| || ||||:|| ||:| ||||:|
Db   1726 SY-SPTSPSY-SPTSPSY-SPTSPNY-SPTSPNYT-PTSP-SYSPTSP-SYSPTSPNY-T 1777

Qy   601 PTTPEELAPTTPEEPPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEP--APT 658
          ||| ||:| ||| |||||:| |||||:| ||:
Db   1778 PTSP-NYSPTSPSY-SPTSPSY-SPTSP-----YTPQSPT-YTPSSPSYSPS 1828

Qy   659 TPKETAPTTPKGTAPTTPLKEPAPTTPK-KPAPKELAPTTTKEPTSTTSDKPAPTTPKGTA 717
          :||| |:| || ||||:| ||:|||||:| |||:||:|:|
Db   1829 SP-SYSPTSPKYT--------PTSPSYSPSSPEYTPTSPK--YSPTSPXYSPTSPK-YS 1875

Qy   718 PTTPKEPAPTTPKEPAPTTPKGTAPTTTLKEPAPTTPKKPAPKELAPTTPKGPT-STTSDK 776
          :|||| |:| :|| ||||:| ||:| ||||:||| :|:|
Db   1876 PTSPTY-SPTTPKY-SPTSP--TYSPTSPVYTPTSPK-----YSPTS--PTYSPTSPK 1922

Qy   777 PAPTTPKETAPTTPKEP--APTTPKEPAPTTPETPPPTTPSEVSTPTTTIHKSPDE 834
          :|||||:| |:|:| ||| |:| :| ||||:|
Db   1923 YSPTSP-TYSPTSPKGTSYSPTSPGY-SPTSP-----TYSLTSPAI-----SPDD 1965

Qy   835 STPE 838
          |::|
Db   1966 SDEE 1969

RESULT 15
NFH_HUMAN

ID   NFH_HUMAN   STANDARD;   PRT;   1020 AA.
AC   P12036;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE   (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN   NEFH OR NFH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88328981; PubMed=3138108;
RA   Lees J.F., Shneidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
RT   "The structure and organization of the human heavy neurofilament
RT   subunit (NF-H) and the gene encoding it.";
RL   EMBO J. 7:1947-1955(1988).
-!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC   AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC   SUBSERVED BY THE TWO SMALLER NF PROTEINS.
-!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC   PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC   THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC   INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC   OF AXONAL CALIBER.
-!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC   OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC   LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC   COINCIDENT WITH A CHANGE IN THE INTERMEDIATE NEUROFILAMENT FUNCTION.
-!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

DR   EMBL; X15306; CAA33366.1; -.
DR   EMBL; X15307; CAA33366.1; JOINED.
DR   EMBL; X15308; CAA33366.1; JOINED.
DR   EMBL; X15309; CAA33366.1; JOINED.

```
DR   PIR; S00979; QFHUH.
DR   MIM; 162230; -.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation.
FT   DOMAIN      1    100       HEAD.
FT   DOMAIN    101    413       ROD.
FT   DOMAIN    101   1020       TAIL.
FT   DOMAIN    101    132       COIL 1A.
FT   DOMAIN    133    145       LINKER 1.
FT   DOMAIN    146    244       COIL 1B.
FT   DOMAIN    245    266       LINKER 12.
FT   DOMAIN    267    288       COIL 2A.
FT   DOMAIN    289    292       LINKER 2.
FT   DOMAIN    293    413       COIL 2B.
SQ   SEQUENCE   1020 AA;  111780 MW;  1177C9DC83DCF1D4 CRC64;
```

Db   958 KEEKAKK-------PEEKPKTEAKAKEDDKTLSKEPSKRPKAEKAEKSSSTDQKDSKPPEKAT 1012
Qy   849 EN 850
         |:
Db 1013 ED 1014

```
Query Match         6.9%;  Score 502; DB 1;  Length 1020;
Best Local Similarity  26.7%;
Matches   209; Conservative   81;  Pred. No. 4.6e-13;
                                    Mismatches  356;  Indels  136;  Gaps  35;

Qy   102 KKTKVIESE--EITEHSVS-ENQESSSSSSSSSSSSTIWKIKSSKNSAANRELQKKLK 158
         | ||  :    : || ::  : : |:
Db   336 KSTKDSLERQSELEDRHQADIASYQEAIQQLDAELRNTKWEM----AAQLREYQDLLN 390

Qy   159 VK-----------------DNKKNRTKKPTPKPPVVDEAGSGLDNGDFKVTPDTSTTQHNK 204
         || :                 :  : :  ||   :|  |  :    ||  ::  :|
Db   391 VKMALDIEIAAYRKLLEGEECRIGFGPIP--------FSLPEGLPKI--PSVST--HIK 437

Qy   205 VSTSPKITTAKPINPRPSLPFNSDISKETISLTVNKETTVETKETTTTNKQTSTDGKE-KT 263
         | :  |:  |:  |: :        :  : |         ::|| || | : ||  : ||
Db   438 VKSEEKIKVVE--------------------------KSEKETVIVEE----QTEETQVTEEVTEEEKEAKE 480

Qy   264 TSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPTTPKEPASTTPK 323
         | |:|    :     |:  | |       ||  :   || ||||  | ||  |||  |||
Db   481 EEGKEEGGGEEEEAEG-------------GEEETKSP-----PAEEAASPEKEAKSPVKE 522

Qy   324 EPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTPKEPAPTT 383
         |  |  | |||| ||||| ||    || |||   |||| ||  || |||| ||||| ||
Db   523 EAKSPAEAAKSPEKEAKSPAEVKS--PEKAKSPA------KEEAKSPPEAKSPEKEAKSPAE 577

Qy   384 TKSAPTTPKEPAPTTTPKKPAPTTPKEPAPTTTPKEPTTTPKEPAPTTTKEPAPTTPKEPAP 443
         |   || ||| || || ||  || ||| | :       ||  | |  ||  ||  ||| |
Db   578 VKS-PEKAKSPAKEEAKSPA---EAKSPEKAKSPVKEEAKSPA-EAKSPVKEEAKSPAE 631

Qy   444 TAPKKPAPTTPKEPAPTTPKEPAPTTTPKEPAPTTTKEPSPTTTPKEPAPTTKSAPTTPKEPASPTTKRAPTA 503
         |   || ||| ||  || ||| || |||  ||  ||  |  || |||  ||  ||||| ||
Db   632 VKSPEKAKSPTKEEAKSPEKEAKSPEKEAAKSPEKAKSPVKAEAKS-PEKAKSPAEAKS 690

Qy   504 APTTPKEPSPTTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKKPAPTA 563
         ||  || || |||||| ||  ||| |||  || |||    |||  |:
Db   691 -PEKAKSP---VKEEAKSPVKEEAKSPVKEEAKSPV---KEEAKTPEKAKSPV- 740

Qy   564 PKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPE---KPAPTTPEEL-APTTPEEPTPTTP 620
         |||| || |||  || || |   |       |      |||  | |  ||  |||| | |:
Db   741 -KEEAKSPEKAKSPEKAKTLDVKSPEAKTPAKEEARSPADKPPEKAKSPVKEEAVKSPEKA 799

Qy   621 EEPAPTTPKAAANTPKEPAPTTP----KEPAPTTPKEPAPTTPKGTAPTTL 676
         : ||  |  || |  ||  |        : ||| |||| ||||| || ||
Db   800 KSPLKADAKAPEKEIPKKEEVKSPVKEEPQEVKKEPPKKAEEEKAPATPK--- 852

Qy   677 KEPAPTTPKKPAPKELAPTTTKEPTSTTTSDKPAPTTPKGTAPTTPKEPAPTT 736
         : || |  |  ||| ||| |||| | |||:    || ||  || ||||| |||
Db   853 TEEKKDSKKEEAPKKAAP----KPKVEEKKEPAVEKPKESKVEAKKEEAEDKKKVP---T 905

Qy   737 PKGTAP--TTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPA-PTTPKETAPTTPKEP 793
         ||  ||       | |:| | | |  || |||  | ||| |   ||  |||  || |||
Db   906 PEKEAPAKVEVKEDA-------KPKEKTEVAKKEPDDAKAKEPSKPAEKKEAAP-EKKDT 957

Qy   794 APTTPKKPAPTTPETPPPTTSEV--STPTTTKEPTTIHKSPDE---STPELSAEPTPKAL 848
         ||                |::    |||  | |||| | ||:     | ||  || || |
Db   958 ...
```

```
                                 GenCore version 4.5
                           Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:              April 26, 2002, 16:25:14 ; Search time 61.21 Seconds
                                              (without alignments)
                                              501.094 Million cell updates/sec

Title:               AA2
Perfect score:       7276
Sequence:            1 MAWKTLPIYLLILLSVFVIQ..........ARAITTRSGQTLSKVWVNCP 1363
Scoring table:       BLOSUM62
                     Gapop 10.0 , Gapext 0.5
Searched:            212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters:      212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Issued_Patents_AA:*
      1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
      2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
      3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
      4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
      5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
      6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result           %
No.   Score  Query    Length  DB   ID                   Description
             Match                                  
-----  -----  -----   ------  --   ---------------      -----------
 1     522    7.2     805     4    US-09-103-429A-4     Sequence 4, Appli
 2     513    7.1     1837    4    US-08-928-361B-5     Sequence 5, Appli
 3     506.5  7.0     744     6    522236-25            Patent No. 5202236
 4     499.5  6.9     1721    3    US-08-700-651-5      Sequence 5, Appli
 5     498.5  6.9     786     4    US-09-103-429A-3     Sequence 3, Appli
 6     498.5  6.9     1721    3    US-08-928-361B-6     Sequence 6, Appli
 7     488.5  6.7     826     1    US-07-638-431-2      Sequence 2, Appli
 8     488.5  6.7     826     5    PCT-US92-00018-2     Sequence 2, Appli
 9     476.5  6.5     652     5    520236-13            Patent No. 5202236
10     452    6.2     960     4    US-09-219-849-5      Sequence 5, Appli
11     425.5  5.8     1185    4    US-09-041-886-23     Sequence 23, Appl
12     424.5  5.8     1867    2    US-08-475-537A-5     Sequence 5, Appli
13     424.5  5.8     1867    2    US-08-475-537A-5     Sequence 5, Appli
14     424.5  5.8     2035    4    US-09-083-116-5      Sequence 5, Appli
15     424.5  5.8     2035    4    US-09-083-116-2      Sequence 2, Appli
16     419.5  5.8     2476    2    US-09-083-116-2      Sequence 2, Appli
17     417    5.7     829     1    US-08-642-255-132    Sequence 132, App
18     417    5.7     829     1    US-08-397-633A-53    Sequence 53, Appl
19     417    5.7     837     1    US-08-175-155-68     Sequence 68, Appl
20     417    5.7     837     1    US-08-477-509B-103   Sequence 103, App
21     417    5.7     837     1    US-08-642-255-101    Sequence 101, Appl
22     417    5.7     837     2    US-08-707-237A-75    Sequence 75, Appl
23     417    5.7     837     3    US-08-482-085B-103   Sequence 103, Appli
24     417    5.7     837     1    US-08-397-633A-50    Sequence 50, Appli
25     413.5  5.7     907     3    US-08-783-774-2      Sequence 2, Appli
26     413.5  5.7     907     5    PCT-US95-0461A-19    Sequence 19, Appli
27     404.5  5.6     408     1    US-07-609-716-65     Sequence 65, Appl
28     404.5  5.6     408     4    US-08-475-411A-65    Sequence 65, Appl
29     404.5  5.6     408     4    US-08-478-029A-65    Sequence 65, App
30     398.5  5.5     682     1    US-08-642-255-126    Sequence 126, App
31     398.5  5.5     682     4    US-08-397-633A-36    Sequence 36, Appl
32     387.5  5.3     1537    1    US-08-325-267A-2     Sequence 2, Appl
33     383.5  5.3     1231    3    US-08-904-263A-4     Sequence 4, Appl
34     381    5.2     1848    4    PCT-US95-10661A-6    Sequence 6, Appli
35     381    5.2     1848    5    US-08-296-791-6      Sequence 6, Appli
36     376    5.2     960     4    US-09-219-849-6      Sequence 6, Appli
37     373    5.1     1786    4    US-09-219-849-6      Sequence 6, Appli
38     371    5.1     761     2    US-08-707-237A-84    Sequence 84, Appl
39     371    5.1     762     1    US-08-397-633A-26    Sequence 26, Appl
40     371    5.1     762     1    US-08-642-255-114    Sequence 114, App
41     370.5  5.1     1064    1    US-08-642-255-62     Sequence 62, Appl
42     370.5  5.1     1187    3    US-08-320-559-28     Sequence 28, Appl
43     370.5  5.1     1187    3    US-08-545-860D-28    Sequence 28, Appli
44     370.5  5.1     1187    5    PCT-US94-04496-28    Sequence 28, Appl
45     370.5  5.1     1210    1    US-08-320-559-26     Sequence 26, Appl
```

RESULT 1
US-09-103-429A-4

; Sequence 4, Application US/0910342 9A
; Patent No. 6187558

; GENERAL INFORMATION:
;   APPLICANT: Granados, Robert R
;   APPLICANT: Wang, Ping
;   TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
;   TITLE OF INVENTION: cDNA and Related Products and Methods
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
;   STREET: 118 No. 6187558th Tioga
;   CITY: Ithaca
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 14850
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/103,429A
;   FILING DATE: 24-JUN-1998
;   CLASSIFICATION: 800
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Michaels, Christopher A
;   REGISTRATION NUMBER: 34,390
;   REFERENCE/DOCKET NUMBER: BTI-39
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (607) 256-2000
;   TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 805 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Trichoplusia ni
;   TISSUE TYPE: peritrophic membrane
US-09-103-429A-4

                              ALIGNMENTS

Query Match   7.2%;   Score 522;   DB 4;   Length 805;
Best Local Similarity   27.8%; Pred. No. 9.2e-26;

```
RESULT 2
US-08-928-361B-5
Sequence 5, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
  APPLICANT: Petersen, Carolyn
  TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
  TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
  TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
  TITLE OF INVENTION: SPECIES INFECTIONS
  NUMBER OF SEQUENCES: 30
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: PETERS, VERNY, JONES & BIKSA
    STREET: 385 Sherman Avenue, Suite 6
    CITY: Palo Alto
    STATE: CA
    COUNTRY: USA
    ZIP: 94306-1840
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/928,361B
    FILING DATE: 12-SEP-1997
    CLASSIFICATION:
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 60/026,062
    FILING DATE: 13-SEP-1996
  ATTORNEY/AGENT INFORMATION:
    NAME: Verny, Hana
    REGISTRATION NUMBER: 30,518
    REFERENCE/DOCKET NUMBER: 480.76-1(HV)
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 650-324-1677
    TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 5:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1837 amino acids
    TYPE: amino acid
    STRANDEDNESS:
    TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-928-361B-5

Query Match  7.1%;  Score 513;  DB 3;  Length 1837;
Best Local Similarity 21.6%;  Pred. No. 9.2e-25;
Matches 307;  Conservative 105;  Mismatches 539;  Indels 470;  Gaps 49;

Qy   67 HNPTSPPSSKKAPP-----PSGASQTIKSTMTKRSPKP--------------------PNKKKTKKV 107
Db   83 HADASTTSSAYSAPFELDVSGV--PIEPNTRRMVDPVSLMLFDNSTGVWYDPN----TNSI 137

Qy  108 IESE--EITEEHSVSENQESSSSSSSSSSSSSSSSTIWKIKSSKNSAANRELQKLKVKDNKKN 165
Db  138 LEGSIAGIRSESCIVSELNFTSTIGFTDISMNNPV--SITSGELKDPNKQAPISGSRSC 195

Qy  166 RTKKKPTPKPPVVDEAGSGLDNG-DFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLP 224
Db  196 GWKQ--------------GYSIDSSTGFRVDS----------IITGLPTDPYPNCP 226

Qy  225 PNSDISKETSLTVNKET-----IVETKETTTTNKQTST-----DGK----EKTTSAKE 268
Db  227 FNPVTGNLIVSRSTGKIIPNTYAGVVRSNETKTTEPSANINFLLVDPKINAPCNSENSFEQ 286

Qy  269 TQSIEKTSAKDLAPISKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTT 328
Db  287 VQIFDMGSKVVIPYTKCVGVKHTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTT 341

Qy  329 TIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTPKEPAPTTPKEPAPTTPKKSAP 388
Db  342 TTTTTTTTT----TTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTT 395

Qy  389 TTPKEPAPTTPKKPAPTTPKEPAPTTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTAPKK 448
Db  396 TTTTTTTTT----TTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTKK 455

Qy  449 PAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKEPAPTTKSAPTTPKEPAPTTPKEPAPTTP 508
Db  456 PTTTT-------TTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTT 509

Qy  509 KEPSPTTIKEPAPTTPKEPAPTTPKK----------------------- 534
Db  510 KKPTTTTTTATTTTTTSETESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRIENGMAFT 569
```

```
Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;

Qy  283 TSKVLAKPTPKPAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAP-----TTP 337
Db    9 TALGIVAARPEVSDAEKNPALHEPHPDXP--PAEQXXLLPXEYDCTKFYYCGEYGLKFIAP 66

Qy  338 KEPAPTT-----------TKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTTK 385
Db   67 RDCAPGTEPKFSAQTCVHAALAGCTLPGPPAETT---QAPATTQ--APTTTT 119

Qy  386 SAPTTPKEPAPTTPKPKPAPTTPKKPAPTTPKEPTPTTPKEPAPTTPKEPAPTA 445
Db  120 QAPTTTTQATTTT-------QAPTTQAPTTT---QAPTTTQAPTTTT------ 156

Qy  446 PKKPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKSAPTTTKEPAPTTTKSAP 505
Db  157 ---QAPTTTTQ-APTTTQ-APTTTQ------------APTTTQ-APTTTQAA 198

Qy  506 TTPKEPSPTTTKEPAPTTP--KEPAPTTPKKPAPTTPKEPAPTTPK------------ 549
Db  199 TTPAATTPAAT-TPAATTPAATTPAATTPCVPAPTS---APVWPPICELLPNGCPADFD 253

Qy  550 ------------EPAPTTTK 557
Db  254 IHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTE 313

Qy  558 K----------PA 560
Db  314 GNEDEDIDIGDLLDNGCPANFEIDWLLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQ 373

Qy  561 -------------PTAP--KEPAPT---------TPKET 575
Db  374 QCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHES 433

Qy  576 -------APTTPKKLTPTT--------PEKLAPTTPEKPAP 601
Db  434 DCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAP 493

Qy  602 TTPEELAPTTPEEPTPTTPEEPAPTTPKEPAPTTPKEPAPTTPK 661
Db  494 TAAPTAAPTAAPTAAPSTVVPPA-TPPATAAPAVPPTTAAIPT---PAPTAAPTTAA 548

Qy  662 ETAPTT----PKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAP 718
Db  549 PESPTTVTVVPPTAAPTA-----APTT---AVPEIPIVTSAPTTAAPT--AAPTAAPTAAP 598

Qy  719 TTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGEPTST-TSDKP 777
Db  599 TTAVPEIPTT----VTSPPTAAPTT-AAPAPNT------------TVTVPPTAAPTTAAP 641

Qy  778 APTTPKEPAPTTPKEPAPTTPKKPAPTTPETPEPPTTSEVSTPTTTKEPTTIIHKSPDESTP 837
Db  642 AP--------NTTVTAAPTTAAPTT-AAPAPNTVTVPPTAAPTAAP-----AP 685

Qy  838 ELSAEPTPKALENSPKEPCGVPTTKTPAATKPE 869
Db  686 NTTTAAP---------VTTTISAPATT-PE 703
```

```
RESULT  3
5202236-25
Patent No. 5202236
APPLICANT: MAUGH, KATHY J.;ANDERSON, DAVID M.;STRAUSBERG,
SUSAN L.;MCCANDLISS, RUSS;WEI, TENA;FILPULA, DAVID
TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
PROTEIN
NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/528,762
FILING DATE: 25-MAY-1990
APPLICATION NUMBER: 82,456
```

```
;   FILING DATE: 07-AUG-1987
;   APPLICATION NUMBER: 933,945
;   FILING DATE: 24-NOV-1986
;   APPLICATION NUMBER: 650,128
;   FILING DATE: 13-SEP-1984
;   SEQ ID NO:25;
;   5202236-25
;   LENGTH: 744

Query Match             7.0%;  Score 506.5; DB 6;  Length 744;
Best Local Similarity   29.9%;  Pred. No. 8.2e-25;
Matches  239;  Conservative  78;  Mismatches 308;  Indels 175;  Gaps 50;
```

```
Qy    535  ------------------------------------------------------------  534
Db    570  MIPNDGTHVRFRFKVKDVGNTISVRCGKGAGKLEPPDRSLDFTIPPVAGHNSCSIIVGVS  629

Qy    535  ------------------------------------------------------------  535
Db    535  ----------------------PAPTTKE---------PAPTTKEPAPTTKK          558

Qy    630  GGGKIHVSPYGSKDVSLISAPIQPCELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTAK

Qy    559  PAPTAPKEPAPTTPKETAPTTPKKLTPTTPKEKLAPTTPEELAPTTPEEPTPPTT      618
Db    689

Qy    690  PTTTT-TGAPGQPTTTTTTGSPSKPTTTTTTTKAFTTT----TILNPIITTTTTQKPTTT  741
Db    619  TPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKE          678

Qy    742  T----TT--KVPGKPPIATTTTTTLKPIVTTTTTTKATTITTTTTVPTT-----TTTTKR  787
Db    679  PAPTTPKKPAPK--------ELAPTTPKEPTS

Qy    719  TTPKEPA----------------PTTPKEPAPTTPK----------------          738
Db    788  DEMTTTTTTPLPDIGDIEITPIPIEKMLDKYTRMIYDYNSGLLLDSNDEPIPGSQACQIAD  847

Db    848  TSNLPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSVSYLAAKNLIVDTDE   907

Qy    739  --GTAPTTLK----EPAPTTPKKPAPKELAPTTI--------KGPTSTTSDKP         777
Db    908  TYGLPIDTLTGYPLDPVSLIPNPEFGELFPDPISDEIMNGTIAGIVSGISASESLLSQKS  967

Qy    778  APTTPKE--------TAPTTPKEPAPTTPKKPRAPTTPETP----PPTTSEVSTPTT     821
Db    968  APIDPAINMVVGEFGGLLNPATGVMIPGSLGPSEQPFSPEIEDGGIIPP---EVAAANA   1024

Qy    822  TKEPTTIHKSPDESTPE-------LSAEPTTPKALENSPKEPGYP                 858
Db   1025  DKFKLSIPPSVPESIPEKDQKIDSISELMYDIESGRLIGQVSKRPIPGSIAGDLN----P  1080

Qy    859  TTKTPAATKPEMTTTAKDKTTERDLRTTPETTTAARKMTKETA---TTTEKTTESKIPA   914
Db   1081  IMKTPQT-----DSVTGKPI--DPTIGLP-FNPPTGHLINPINPNNNNMDSSPAG      1126

Qy    915  TTIQVTSTTTQDTTPFKITTLKTTTTLAPK---VTTTKKITTTTEIMNKPEETAKPKDRAT- 971
Db   1127  AYKPAVSNGIKTDNVYGLPVDEITGL-PKDPVSDIPFNSTTGELVD---PSTGKPINNYTA  1183

Qy    972  ---------NSKATTPKPQKPTKAPKKPETSTKKPKTMPRVRKPKTTPTPRKMTS       1016
Db   1184  GIVSGKRGLPPIEDENGLPDPSTKLPIDGNGNLVNPETNSTVSGSTSGGSTKPKGIPVN   1243

Qy   1017  ---TMPE-------LNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSEDAGG          1058
Db   1244  GGGVVPDEEAKDQADKGKDGLIVPPTNSTLNKDPVTNIQYSNTGNI----INP--ETGKV  1297

Qy   1059  AEGETPHMLLRPHVEMPEVTPD-----MDYLPRVPNQGIIINP----------         1096
Db   1298  IPGSLPGSLNYPSFNTPQOTDEITGKPVDTVIGLPYDPSTGEIIDPATKLPIPGSVAGDE  1357

Qy   1097  MLSDETNICNGK------PVDGLTTLRNGTLV                              1122
Db   1358  ILTEVLNITTDEVTGLPIDLETGLPRDPVSGLPQLPNGTLV                    1398
```

```
Qy    209  PKITTAKPINPRPSLPPNSDTDSKETSLTVNKEITVVETKEITTINKQTSTDGKEKTTSAKE  268
Db     24  PKMTYPPTYKPKPSYP---------TYKSKPTY-----KPKIT-----               53

Qy    269  TQSIEKTSAKDLAPTSKVLAKPTTPKAETTTKGPALTTPKEPTPTTPKE---PASTTPKE   324
Db     54  YPPTYK-AKPS----YPPTYKPKKTYPPTYKPKLTYPPTYKPKP                   92

Qy    325  PTPTTIKSAPTT-TPKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA--P  381
Db     93  SYPPTYKSKPTYKPKITYPPTYKAKPSYPPTYKPKKTYPPTYKPKASYP             151

Qy    382  TTTTKSAPTTPKEPAPTTPKKPAPT-TPKEPAPTT-------PKEPTPTTPKEPAPTTKEP  433
Db    152  PTYKPKPSYP--PSYKTKKTYPPTYKPKLTYPPTYKPKP                        209

Qy    434  -APTTPKEPA--PTTPKEPA---PTAPKEPAPTTPKETAPTTPKKLTPTT             482
Db    210  YPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYP                     269

Qy    483  -PTTTKEPAPTTPKEPSPTTTKEPA--PTTPKEPA----PTTPKKP                 535
Db    270  YPSTYKAKPSYT---YPSTYKAKPTYP--PTYKAKPSYPPTYKAKP                 322

Qy    536  A--PTTPKEPA---PTTPKKPA-PTAPKEPAPTTPKEPAPTTPKETAPTTPKKLTPTT     587
Db    323  SYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYPS         382

Qy    588  PEKLAPTTPE--KPAPTTPEELAPTTPEEPT--PTTPEEPA--PTTPKA--AAPNT----P  636
Db    383  TYKAKPSYPPTYKAKPSYP----PTYKAKPSYPPTYKAKPSYPPTYKAKPTYP          438

Qy    637  KEPAPTTPKEPAPTT----PKEPAPTTPKETAPTT-----PKGTAPTTLKEPAPTTPKKPAPK 690
Db    439  SYPPTYKPKISYPPTYKAKPSYPSTYKAKPSYPPTYKAKPSYPSTYKAKSSYPPTYKAKP   498

Qy    691  ELAPTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPA-   749
Db    499  SYPPTYKAKPTYPSTYKAKPTYK---ARPSYP-----YPPTYKAKPSY              546

Qy    750  -PTTPKKPAPKELAPTTTKGPT---STTSDKPAPTTPKETAPTTPKEPAPTTPKKPA--PT  804
Db    547  PPTYKAKPTYK-AKPTYKAKPSYSPSTYKAKPSYSPTYKAKPSYP----PTYKAKPSYPPT  601

Qy    805  ---TPETPPPTTSEVSTPTTTKE----PTTIHKSPD-----ESTPELSAEPT-PKALENS  851
Db    602  YKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYPSTYKAK         661

Qy    852  PKEPGVPTTKPAA--------TKPEMTTTAKDKTTERDLRTTPETTTAAP------KMT    897
Db    662  PSYP--PTYKAKPSYPPTYKAKPTYKAKPTYPSTYKAKPTYPSTYKAKP              719

Qy    898  KETATTTEKTTESKITATTT                                           917
Db    720  YKAKPTYPSTYKAKPTYPST                                           739
```

```
RESULT  4
US-08-700-651-5
```

```
Sequence 5, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
 APPLICANT: PETERSEN, CAROLYN
 APPLICANT: LEECH, JAMES
 APPLICANT: NELSON, RICHARD, C.
 APPLICANT: GUT, JIRI
 TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
 TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
 TITLE OF INVENTION: INFECTIONS
 FILE REFERENCE: 480.19-4(HV)
 CURRENT APPLICATION NUMBER: US/08/700,651B
 CURRENT FILING DATE: 1997-08-14
 EARLIER APPLICATION NUMBER: 08/415,751
 EARLIER FILING DATE: 1995-04-03
 NUMBER OF SEQ ID NOS: 15
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 5
  LENGTH: 1721
  TYPE: PRT
  ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match              6.9%;  Score 499.5; DB 3;  Length 1721;
Best Local Similarity   21.8%;  Pred. No. 6.2e-24;
Matches 300;  Conservative 94;  Mismatches 510;  Indels 469;  Gaps 44;

Qy   90 STTKRSPKPPNKKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSSSSSSSSSTIWKIKS------ 143
Db   38 SITSGERKDPNKQAT------------------ISGSRSCGWKQGYSIDSSTGFRVDSITGLPT 83

Qy  144 --SKNSAANRELQKKLKVKDNKKNRIKKKPPPKPPVVDEAGSGLDNGDFKVTTPDISTTQ 201
Db   84 DPYSNCPFN-------PVTGNLVSRSTGKTIPN-----TYAGVYRSN-ETKITEPSANT-- 129

Qy  202 HNKVSTSPKITTTAKP-----------INPRPSLPPNSDISKE------------TSLTVN 238
Db  130 YAGVYRSNETKTTEPSANTNFLLVDPKINAPCNSENSPEQGQIFDMGSKVYIPYTKCVGV 189

Qy  239 KETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTT 298
Db  190 KHTTTTTTTTTTTTTTTTTTT-----------------------TTTT 211

Qy  299 KGPALTTPKEPTPTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPT 358
Db  212 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 271

Qy  359 TTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTIKSAPTTPKEPAPTTPKEPAPTTPKEP 418
Db  272 TTTTTTTTTTTTTTTTTTTTTTTKPTTTTTTTTTTTTTTTTTTTTTTTT 331

Qy  419 TPPTTPKEPAPTTPKEPAPTTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTTP 478
Db  332 TTTTTKKPTTTTTTTTTTTT----------TTTTTTTTTTTTTTT 377

Qy  479 KEPAPTTTKSAPTTTKEPAPTTTKSAPTTPKE--------- 510
Db  378 TKPTTTTTTTTTTTTKPTTTTTTTTTTTTTSETESVIKPDEWCWLEKNGECEAKGAIYVG 437

Qy  511 ------------PSPTT--------------TKE 518
Db  438 VIGKGDGRIENGMAFTMIPNDDTHVRFRFKYKDVGNTISVRCRKGAGKLEFPDRSLDFTIP 497

Qy  519 P----------------APTTPKEPAPTTPKKPAPTTPKE------ 542
Db  498 PVAGHNSCSIIVGVSGDGKIHVSPYGSKDVSLIS----APIQPSELFNEVYCDTCTAKYG 553

Qy  543 ---PAPTTPKEPAPTTKKPAPTAPKEPAPTTPKKLTPTTPEKLAPTTPEKP 599
Db  554 AIHSGYQTSADEVTTTTTTAKPTTTT--TGAPGQPTTTTGSPSKPTTTTKATTTT---- 607

Qy  600 APTTPEELAPTTPEEPPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTT 659

Db  608 --TTLNPIIITTTTQKPTTTT---------TT---KVPGKPIATTTTTLKPILVTTTTTKATTT 656
Qy  660 PKETAPTTPKGTAPTTLKEPAPTTPEKPAPK-------ELAPTTTKEPTS----------- 702
Db  657 TTTTVPTT----TTTTKRDEMTTTTTPLPDIGDIEITTPIEKMLDKYTRMIYDYNSGL 711

Qy  703 --TTSDKPAPTTPKG-TAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKGTAPTTLKE---PAPTTPKP 756
Db  712 LLDSNDEPIPGSQAGQIADTSNLFPVQTHKSTGLPIDPMVGLPFDPEKSGNLVHPYTNQIM 771

Qy  757 APKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEV 816
Db  772 SGLSVSVIAAKNLTVDTDETYG--LPIDTLTGYPLDPVSLLPFN--PETGELFDPISDEI 827

Qy  817 STPTT----------TTKEPTTIHKS-------------------------PDEST- 836
Db  828 MNGTIAGIVSGISASESLLSQKSALIDPATNMVVGEFGGGLLNPATGVMIPGFLGPSEQTQ 887

Qy  837 ---PELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTER--- 881
Db  888 FSPEIEDGGIIPPEVAAANADKFKLSIP--PSVP------ESIPE-----KDQKIDSISE 934

Qy  882 ------------DLRTTPETTTAAPKMTKETATTTTEKTTESKITATT-TQVTSTTTQD 926
Db  935 LMYDIESGRLIGQVSKRPITPGSIAGDLNPIMKIPTQTDSVTGKPIDPTTGLPFNPPTGHL 994

Qy  927 TTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKP--EETAKPKD------RATNSKATT 977
Db  995 INPTNNNTMDSSFAGAYKYAVSNGIKTDNVYGLPVGEITGLPKDPGGSDIPFNSTTGELVD 1054

Qy  978 PKPQKPTKAPKKPTSTKKPKTMP-------------------RVRKPKTTIPTTPRKMTS 1016
Db 1055 PSTGKPINNSTIAGIVSGKPGLPPIEDENGNLFDPSTNLPIDGNNQLVNPETNSTSVSGSTS 1114

Qy 1017 ---TMPE------------LNPTSRIAEAMLQTTTRPNQTPNSKL 1046
Db 1115 GTTKPRPGIPVNGGGVPDEEAAKDGKDGLIVPPINSINKDPVTNYGYSNTTGNI-- 1172

Qy 1047 VEVNPKSEDAGGAEGGETPHMLLRPHVFMPEVTPD--------MDYLPRVPNQGIIINP-- 1096
Db 1173 -INP-ETGKVIPGSLPGSLNYPSFNTPQQTDEITGKPVDTVTGLPYDPSTGEIIDPAT 1228

Qy 1097 --------------MLSDETNICNGK--------------PVDGLTTLRNGTLV 1122
Db 1229 KLPIPGSVAGDEILTEVLNITTTDEVTGLPIDLETGLPRDPVSGLPLPNGTLV 1281


RESULT 5
US-09-103-429A-3
Sequence 3, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
 APPLICANT: Granados, Robert R
 APPLICANT: Wang, Ping
 TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
 TITLE OF INVENTION: cDNA and Related Products and Methods
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
  ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
  STREET: 118 No. 6187558th Tioga
  CITY: Ithaca
  STATE: NY
  COUNTRY: USA
  ZIP: 14850
 COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/103,429A
  FILING DATE: 24-JUN-1998
```

```
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
  NAME: Michaels, Christopher A
  REGISTRATION NUMBER: 34,390
  REFERENCE/DOCKET NUMBER: BTI-39
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (607) 256-2000
  TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 786 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: protein
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
  ORGANISM: Trichoplusia ni
  TISSUE TYPE: peritrophic membrane
US-09-103-429A-3
```

```
Query Match      6.9%;  Score 498.5;  DB 4;  Length 786;
Best Local Similarity   29.1%;  Pred. No. 2.8e-24;
Matches 200;  Conservative 25;  Mismatches 207;  Indels 255;  Gaps 32;

Qy  290  PTPKAETTTKGPALTTPKEPPTTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTKSAP  349
Db   93  PGPPAE-TTQAPATT----QAPTTQAPTTT------QAPTTTQ-ATTTTQAP  135

Qy  350  TTPKEPAPTTTPKEPAPTTTPKEPPTTTPKEPASTTPKEPAPTTPKKPAPTTPKE  409
Db  136  TTTTQ--APTTTQ-ATTT--QAPTTTQ-APTTTTQAPTTTQ-APTTTQ-  183

Qy  410  PAPTTPKEPTPTTPKEPAPTTTPKEPAPTTPKEPAPTTPKEPAPTT-PKEPA--  466
Db  184  -APTTQAPTTITQAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC  241

Qy  467  -------------------------PTTTKEPSPTTPK-  479
Db  242  ELLPNGCPADFDIHLLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPVYQRCDSPANVEC  301

Qy  480  ---EPAPTTT---  486
Db  302  DGEISPAPVVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR  361

Qy  487  ------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTPKEP  519
Db  362  CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP  420

Qy  520  APTT-------PKEP---------------APTTPKKPAPTT-------  539
Db  421  ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD  480

Qy  540  ---PKEPAPTTPKEPAPTTTKKPAPTTAPKEPAPTTPKETATTAPTTPKKLITPTTPEKLAPTTP  596
Db  481  NQCTSTAPTTAAPTTAAPTTAAPTTAA--APSTVVPPATTPAATAAPVPPTT--AIPT--  533

Qy  597  EKPAPTTPEELAPTTTPEEPPTTPTTPEEPAPTTPKAAAPNT--PKEPAPTTTPKEPAPTTPKE  654
Db  534  --PAPTAAPTTAAPTTAAPESPTTTVTVP-PTAAPTAAPTTAVPEIPITVT--SAPTAAPT  587

Qy  655  PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPAPRELAPTTTKEPTSTTRSDKPAPTTPK  714
Db  588  AAPTAAPTTAAPTTAVPEIPTTVTVSPPTAAPTTAAP--APNTT------  632

Qy  715  GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTTPKKPAPKELAPTTTKGPISTTS  774
Db  633  TAAPTT-AAPAPNTTVVPPTAAPTAAPPVAH----  677

Qy  775  DKPAPTTPKEPTAPTTPKEPAPTTPKKP  801
```

```
Db  678  ---APATTPEDDIDP--PLPNDPINP  699
```

```
RESULT 6
US-08-928-361B-6
Sequence 6, Application US/0892836lB
Patent No. 6071518
GENERAL INFORMATION:
  APPLICANT: Petersen, Carolyn
  TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
  TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
  TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
  TITLE OF INVENTION: SPECIES INFECTIONS
  NUMBER OF SEQUENCES: 30
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: PETERS, VERNY, JONES & BIKSA
  STREET: 385 Sherman Avenue, Suite 6
  CITY: Palo Alto
  STATE: CA
  COUNTRY: USA
  ZIP: 94306-1840
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/928,361B
  FILING DATE: 12-SEP-1997
  CLASSIFICATION:
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 60/026,062
  FILING DATE: 13-SEP-1996
  ATTORNEY/AGENT INFORMATION:
  NAME: Verny, Hana
  REGISTRATION NUMBER: 30,518
  REFERENCE/DOCKET NUMBER: 480.76-1(HV)
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 650-324-1677
  TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 6:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 1721 amino acids
  TYPE: amino acid
  STRANDEDNESS:
  TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-928-361B-6
```

```
Query Match      6.9%;  Score 498.5;  DB 3;  Length 1721;
Best Local Similarity   21.8%;  Pred. No. 7.2e-24;
Matches 300;  Conservative 94;  Mismatches 510;  Indels 469;  Gaps 44;

Qy   90  STTKRSPKPPNKKKTKKVIESEETEEHSVSENQESSSSSSSSSSSSSSTTWKIKS------  143
Db   38  SITSGELKDPNKQAT-----------------ISGSRSCGWKQGYSIDSSTGFRVDSITGLPT  83

Qy  144  --SKNSAANRELQKRLKLKVKDNKKNRTKKPTPKPPVVDEAGSGLDNGDEKVTTPDTSTTQ  201
Db   84  DPYSNCPFN------PVTGNLVSRSTGKTIPN----TYAGVYRSN-ETKTEPSANT--  129

Qy  202  HNKVSTSPKITTAKP--------INPRPSLPPNSDTSKE------------TSLTVN  238
Db  130  YAGVYRSNETKTEPSANTNFLLVDPKINAPCNSENSFQGQIFDMGSKVYIPYTKCVGV  189

Qy  239  KETTVETKETTTNKQTSTDGKERTTSAKETQSIEKTSAKDLAPTSKVLAKPTKAEFTT  298
Db  190  KHTTTTTTTTTTTTTTTTT--------TTT  211

Qy  299  KGPALTTPKEPTPTTPKEPASTTPKEPAPTTTIKSAPTTPKEPAPTTTPKEPAPT  358
Db  212  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  271
```

```
RESULT 7
US-07-638-431-2
Sequence 2, Application US/07638431
Patent No. 5198535
GENERAL INFORMATION:
  APPLICANT: Hoffman, Stephen L.
  APPLICANT: Charoenvit, Yupin
  APPLICANT: Hedstrom, Richard
  APPLICANT: Khusmith, Srisin
  APPLICANT: Rogers IV, William O.
  TITLE OF INVENTION: Protective malaria sporozoite surface protein
  TITLE OF INVENTION: immunogen and gene
  NUMBER OF SEQUENCES: 2
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: A. David Spevack
    STREET: NMRDC Building 1 T-12 National Naval
    STREET: Medical Center
    CITY: Bethesda
    STATE: MD
    COUNTRY: USA
    ZIP: 20814-5044
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.24
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/07/638,431
    FILING DATE: 19910110
    CLASSIFICATION: 424
  ATTORNEY/AGENT INFORMATION:
    NAME: Spevack, Avrom D.
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (301) 295-6759
    TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 826 amino acids
    TYPE: AMINO ACID
    TOPOLOGY: linear
  MOLECULE TYPE: protein
US-07-638-431-2

Query Match          6.7%;  Score 488.5;  DB 1;  Length 826;
Best Local Similarity 26.4%;  Pred. No. 1.3e-23;
Matches 153;  Conservative 54;  Mismatches 223;  Indels 149;  Gaps 34;
```

```
Qy  326  TPTTIKSAPTTPKEPA-PTTTKSAPTTPKEP----APTTPKEP--APTTPKE      378
         :|          |   | |  |   | | |    | | | | |  | | | |
Db  273  TPCKVRDCPQIPIPPVIPNKIPEKPSNPEEPVNPNDPNDPENPNPNNPNPNPNNPNPNPNN  332

Qy  379  PAPTTTPKSAPTTPKEP-APTTPKKP-APTTPKEP-TPTTPKEPAPTTPKEPA      434
         | |   | | | | |  | |  |   | | | |  | | | | | | | | |     |
Db  333  PNNPNPNPNNPNNPNPNNPNNPNPNNPNPNNPNPNNPNPNNPNPNNPNNPHPK       392

Qy  435  PTTTPKEPAPTAPK--KPAPTTPKEPAPTTPKEPSPTTPKEPAPTTTPKSAPTT     492
         :    |   | |    |  | |   | | | |   | |   | |       :
Db  393  RRNPKRRNPNPKPKPKPNKPNPNEPSNPNKPNPN----EPSNPNKPNP---NEPSN     443

Qy  493  TKEPAPTTPKEPSPTTTPKEPAPTTPKEPA-PTTPKKP-APTTPKEPAPTTTPKE     550
         |    | |         | |  | | |   | |     | | | | |
Db  444  PNKPNP---NEPSNPNKPNPN----EPLNPNEPSNPNEPSNPNAPSNPNE--PSNPNE   492

Qy  551  PAPTTTPKKPAPTAPKEP-APTTPKETAPTTPKKLIPTTPEKLAPTTPEELAP      609
         | |   |  | |   |   |  |    | |                | | |
Db  493  PS------NPNEPSNPNEPSNPNE---------PSNPKK------P            523

Qy  610  TTPEEPTPTTPEEPAPTTPKAAAPNTPKEP-APTTPKEPAPTTPKETAPTTP      668
              | |     |  | | |          |  | | | | | | | |
Db  524  SNPNE--PLNPNEP-----SNPNEPSNPEPSNPEE--PSNPKE--PSNPNE-----    564
```

```
Qy  359  TPTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEP      418
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  272  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT       331

Qy  419  TPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPKKPAPTTPKEPAPTTPKEPSPTTP  478
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  332  TTTTTTKKPTTTTTTTTTTTT                                            377

Qy  479  KEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKE                              510
         :|  |||:: ||||||||||||||||||||
Db  378  TKPTTTTTTTTTTTTKKPTTTTATTTTTTSETESVIKPDEWCLEKNGECEAKGATYVG     437

Qy  511  ----------------PSPTT----------                               518
                         ||||:
Db  438  VIGKDGRIENGMAFTMIPNDDTHVRFRKVKDVGNTISVRCRGAGKLEFPDRSLDFTIP     497

Qy  519  P------------------------TKE                                  542
         :                       |||
Db  498  -----APTTPKEPAPTTPKKPAPTTPKE                                  553

Qy  543  ---PAPTTPKEPAPTTPKPAPTAPKPAPTTPKETAPTTPKKLIPTTPEKLAPTTPKEKP    599
            :|:||||||||||||||||||||||||||||||||||||||||||||||||||
Db  554  AIHSGYQTSADFVTTTTAKPTTTT---TGAPGQPTTTTTTGSPSKPTTTTTTTKATTTT    607

Qy  600  APTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAAPNTPKEPAPTTPKEPAPTT         659
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  608  --TTLNPIITTTTTTQKPTTTTT----TT----KVPGKPTIATTTTTLKPIVTTTTTKATTT  656

Qy  757  APKELAPTTTKGPTSTTTSDKPAPTTPKETAPTTPKEPAPTTPKKKPAPTTPTPTPPTTSEV  816
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  712  LLDSNDEPIPGSQAGQIADTSNLFPVQTHKSTGLPIPDPMVGLPFDPFPKSGNLVHPYTNQTM  771

Qy  703  --TTSDKPAPTTPKG-TAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKE---PAPTTPKKP    756
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db  657  TTTTTVPTT-----TTTTKRDEMTTTTTTPLPDIGDIEITPIPIEKMLDKYTRMIYDYNSGL  711

Qy  660  PKETAPTTPKGTAPTTLKEPAPTTPKKPAPK-----ELAPTTKEPTS----------     702
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  608  --TTLNPIITTTTTTQKPTTTTT-----TT-----KVPGKPIATTTTTLKPIVTTTTTKATTT  656

Qy  817  STPTT------------TTKEPTTIHKS------                            836
              :              ||||||||:
Db  772  SGLSVSILAAKNLIVDIDETYG--LPIDILTGYPLDPVSLIPFN--PEIGELFDPISDEI    827

Qy  828  MNGTIAGIVSGISASESLLSOKSALIDPATNMVVGEFGGLLNPAIGVMIPGFLGPSEQTQ   887
         :                                               :
Db  817  ----TTKEPTTIHKS----                                       836

Qy  837  -----PELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTER---       881
              |:||                   : |  |||:
Db  881  APTTPKKP                                                  756

Qy  888  FSPEIEDGGIIPPEVAANADKFKLSIP--PSVP----ESIPE----KDQKIDSISE      934
             :   :                    |::|      : |
Db  934  ...                                                       ...

Qy  882  -----DLRTTPETTTAAPKMTKETATTTEKTTEESKIIFATT-TQVTSTTTQD       926
Db  935  LMYDIESGRLIGQVSKRPIPGSIAGDLNPIMKTPTQTDSVGKPIDPTTGLPFNPPTGHL    994

Qy  935  LMYDIESGRLIGQVSKRPIPGSIAGDLNPIMKTPTQTDSVGKPIDPTTGLPFNPPTGHL   994
Db  927  TTPFKITTLKTTTLAPKVTTTIKTITTTEIMNKP----RATNSKATT              977

Qy  978  PKPQKPTKAPKPTTKPKTMP-------                                   995
         | |  | | | : | |
Db  995  INPINNNTMDSSFAGAYXYAVSNGIKTDNVYGLPVGEITGLPKDPGSDIPFNSTTGELVD  1054

Qy  1017 --TMPE----------                                              1055
Db  1055 PSTGKPINNSTAGIVSGKRPGLPPIEDENGNLFDPSTNLPIDGNNOLVNPEINSTVSGSTS 1114

Qy  1047 VEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPD--------                    1096
Db  1047 GTTKPKPGIPVNGGGVVPDEEAKDQADKGKDGLIVPPINSINKDPVTNTQYSNTTGNI-- 1172

Qy  1115 GTTKPKPGIPVNGGGVVPDEEAKDQADKGKDGLIVPPINSINKDPVTNTQYSNTTGNI-- 1172
Db  1173 -INP-----EIGKVIPGSLPGSLNYPSFNTPQQTDEITGKPVDTTVGLPYDPDSTGEIIDPAT 1228

Qy  1173 -INP-----EIGKVIPGSLPGSLNYPSFNTPQQTDEITGKPVDTTVGLPYDPDSTGEIIDPAT 1228
Db  1097 -------MDYLLRVPNQGIIINP-                                      1096

Qy  1097 -------MLSDETNICNGK---                                        1122
Db  1229 KLPIPGSVAGDEILLTEVLNITTDEVTGLPIDLETGLPRDPVSGLPQOLPNGTLV       1281

Qy  1229 KLPIPGSVAGDEILLTEVLNITTDEVTGLPIDLETGLPRDPVSGLPQOLPNGTLV       1281
Db  524  SNPNE--PLNPNEP-----                                           564
```

```
RESULT  8
PCT-US92-00018-2
Sequence 2, Application PC/TUS9200018
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khusmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: Immunogen and gene encoding
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMRDC Building 1 T-12 National Naval
; STREET: Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 424
; FILING DATE: 19920103
; CLASSIFICATION: PCT/US92/00018
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-4033
; TELEFAX: (301) 295-6759
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein

Query Match          6.7%;  Score 488.5; DB 5;  Length 826;
Best Local Similarity  26.4%;  Pred. No. 1.3e-23;
Matches  153;  Conservative  54;  Mismatches  223;  Indels  149;  Gaps  34;

Qy    669 KGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTT 728
              |||      |     |   |     |    |   |||     ||        |
Db    565 ---------------PSNPEEPNPEE--PSNPKEP----SNPEEPINPEELNPKEPSNPEESN 606

Qy    729 PKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPT 788
              |||||          |||  |   ||||
Db    607 PKEPI-------------NPEDNENPLIIQODEPIEPRNDSNVIPI 649

Qy    789 TPKEP------APTTPKKPAPTTPETPPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAE 842
              |||           | |   |    |     |    |||
Db    650 LPIIPQKGNNIPSNLPENPSDSEVEYPRRNDNGENSNNTMKSKKNI---PNEPIPSPGDN 706

Qy    843 PTPKALENSPK---------- 861
              |  ||
Db    707 PYKGHEERIPKPHRSNDDYVYDNNVNKNNKDEPEIPNNE 745
```

```
RESULT  9
5202236-13
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO:13:
; LENGTH: 652
5202236-13

Query Match          6.5%;  Score 476.5; DB 6;  Length 652;
Best Local Similarity  29.7%;  Pred. No. 5.7e-23;
Matches  221;  Conservative  71;  Mismatches  284;  Indels  169;  Gaps  45;

Qy    209 PKITTAKPINPRPSLPPNSDDTPSKEPSLTVNKEPTVETKEPTTTNKQPTSTDGKEKTTSAKE 268
              ||:|  |    |  | :  |      |    |    |   |
Db     11 PKMTYPPTYKPKPSYPP--------------------TYKSKPTY-----KPKIT---- 40

Qy    269 TQSIEKTSAKDLAPTSKVLAKPTPKKAETTTKGPALTTVKPKEPTPTTPKE----PASTTPKE 324
              :|| |  |          |||           |             |||
Db     41 -------------YPPTYK--AKPS------YPPTYKPKKTYPPTYKPKLTYPPTYKPKP 79

Qy    325 PTPTTIKSAPT-TPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPA--P 381
              |||||||:|      |   |||  :||   | :|   |||   ||
Db     80 SYPPTYKSKPTYKPKITYPPTYKAKPSYPPTYKPKKTYPPT-YKPKLTYPPTYKPKASYP 138

Qy    382 TTTIKSAPTTPKEPAPTTPKKPAPT-TPKEPAPTT----PKEPTTTPKEPAPTTKEPAPTT 436
              ||:|||||||||||||||  ||| :||     ||    |||  |||:||||||||
Db    139 PTYKPKPSYP--PSYKTKKTYPPTYKPKLTYPPTYKPKPKLTYPPTYKPKPSYPPSYKPNKTYPPTYKPKLT 196
```

```
Qy    393 RRNPKRRNPENKPKPNKPNPNKPNPNPNPEPSNPNKPNPN---- 443
              ||       |       |   |:  |      ||  |
Db    393 RRNPKRRNPENKPKPNKPNPNKPNPNPNPEPSNPNKPNP---- EPSNPNKPNP 443

Qy    493 TKEPAPTTKSAPTTPKEPSPTTTKEPAPTTPKEPA-PTTTPKKP-APTTPKEPAPTTPKE 550
              |||  |: ||  |  :|||       | |  |||
Db    444 PNKPNP-----NEPSNPKPNPN-----EPLNPNEPSNPNEPSNPNE--PSNPNE 492

Qy    551 PAPTTPKKPAPTAPKEP--APTTPKETAPTTPKKLTPTPTPEKLAPTTPEKPAPTTPEELAP 609
              | |||   |  :|||      |  |  | |:|   ||  |     |   ||  || ||
Db    493 PS------NPNEPSNPNEPSNPNE---PSNPKK-------------------PSNPNE--P 523

Qy    610 TTPEEPTPTTTEEPAPTTPKAAAAPNTPKEP--APTTPKEPAPTTPKETAPTTP 668
              ||  ||    ||||||||| |     :|||      |  |   ||  || ||
Db    524 SNPNE----PLNPNEP-----------SNPNEPSNPNEPSNPNE---PSNPKE------- 564
```

```
Qy    669 KGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKEPAPTT 728
              |||      |     |   |     |    |   |||     ||        |
Db    565 ---------------PSNPEEPNPEE--PSNPKEP----SNPEEPINPEELNPKEPSNPEESN 606

Qy    729 PKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPT 788
              |||||          |||  |   ||||
Db    607 PKEPI-------------NPEDNENPLIIQODEPIEPRNDSNVIPI 649

Qy    789 TPKEP------APTTPKKPAPTTPETPPPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAE 842
              |||           | |   |    |     |    |||
Db    650 LPIIPQKGNNIPSNLPENPSDSEVEYPRRNDNGENSNNTMKSKKNI---PNEPIPSPGDN 706

Qy    843 PTPKALENSPK---------- 861
              |  ||
Db    707 PYKGHEERIPKPHRSNDDYVYDNNVNKNNKDEPEIPNNE 745
```

```
RESULT 10
US-09-219-849-5
  Sequence 5, Application US/09219849
  Patent No. 6150081
GENERAL INFORMATION:
  APPLICANT:  VAN HEERDE, GEORGE V.
  APPLICANT:  VAN RIJN, ALEXIS C.
  APPLICANT:  BOUWSTRA, JAN B.
  APPLICANT:  DE WOLF, FREDERIK A.
  APPLICANT:  MOOROEK, ANDREAS
  APPLICANT:  WERTEN, MARC W.T.
  APPLICANT:  WIND, RICHELE D.
  APPLICANT:  VAN DEN BOSCH, TANJA J.
  TITLE OF INVENTION:  SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
  TITLE OF INVENTION:  SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
  TITLE OF INVENTION:  PREPARATION THEREOF
  FILE REFERENCE:  2728-2
  CURRENT APPLICATION NUMBER:  US/09/219,849
  CURRENT FILING DATE:  1998-12-23
  NUMBER OF SEQ ID NOS:  50
  SOFTWARE:  PatentIn Ver. 2.1
  SEQ ID NO 5
  LENGTH:  960
  TYPE:  PRT
  ORGANISM:  Artificial Sequence
  FEATURE:
  OTHER INFORMATION:  Description of Artificial Sequence: Illustrative
  OTHER INFORMATION:  amino acid sequence
US-09-219-849-5

Query Match     6.28;  Score 452;  DB 4;  Length 960;
Best Local Similarity  25.6%;  Pred. No. 3.3e-21;
Matches 173;  Conservative  60;  Mismatches 309;  Indels 134;  Gaps  41;
```

```
Qy    437  TPKEPAPTAPKKPA--PTTPKEPA--PTTPKEPAPTTPKEPSPTTP----KEPAPTTTK   487
                :|:         |||:   :|:|:|  ||  :  |:  |||  ||        ||||||:
Db    197  YP---PTYKAKPSYPPTYKAKPSYPPTYKAKPSYPSLIKAKPSYPPTYK          252

Qy    488  SAPTTTKEPA-------PTTTKSAPTTPKEPSPTTTKEPA--PTTPKEPA---PTTPKK   534
                :||   |||          |:  :|  ||  ||  |:  |||  |:   |||  |||
Db    253  AKPYIYKAKPTYKAKPTYPSTYKAKPTYP---PTYKAKPSYPPTYKAKPTYKAKPTYKAK   308

Qy    535  PA---PTTPKEPA--PTTPKEPAPTTAPKEPAPTTPKETAPTTPKKLTPTTPPEK   590
                |:   |||  ||   |||  ||  |  ||  |||  |::|  ||:  |||::|
Db    309  PSYPPTYKAKPSYPPTYKAKPSYP---PTYKAKPSYPPTYKVKPT--YK         362

Qy    591  LAPTTPE--KPAPTTPEELAPTTPEEPT--PTTPEEPA--PTTPKAAAPNTPKEPAPTTPK   645
                |||  |   |||  ||  :|:|  ||  ::  ||   |||   :       ||  |||  |
Db    363  AKPTYPSTYKAKPSYP---PTYKAKPSYPPTYKAKPSYP---PTYKA          413

Qy    646  EPA-PTT----PKEPAPTTPKETAPTTP---PKGTAPTTPKEPAPTTPKEKPAPKELAPTTK   698
                |||  |||     ||  |||  ||  |||     :||  |||  ||  |||   |:||:|||
Db    414  KPTYPSTYKAKPSYPPTYKPKISYPPTYKAKSSYPPTYKAKPSYPPTYKA   473

Qy    699  EPTSTTPSDKPAPTTPKGTAPTTPKEPAPTTPKEPA--PTTPKETAPTTP---PTYKAKP   756
                ||  :: |:  |||  |||  |||  :||  |||  |||   ||  ||  ||:   :||||||
Db    474  KPTYPSTYKAKPTYK--AKPSYP----PTYKAKPS          521

Qy    757  APKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPA--PTTPKKPA--PTTPKKPA   810
                |||  |||||||  ::|  |||  ||  ||  |||  |||   |||||  |  ||:||
Db    522  TYK-AKPTYKAKPSYPPTYKAKPSYP---PTYKAKPSYPPTYPSTYKAKPSYP   576

Qy    810  PPTTSEVSTPPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPPTTKTPAATKPE   869
                |||       ||  :|   ||  ||:   ::|  ||  ||   |  :|  |||       |:
Db    577  PTYKAKPSYPPT------YKAKPSYPPTYKAKPTYKA          620

Qy    870  MTTTAKDKTTERDLRTTPETTTAAP   894
                ||:         |:  ||::|  ||
Db    621  YPPTHKAKPTYKAKPSYPPTHKAKP   645
```

```
Qy    290  PTPKAETTTKGPALTTPKEPTPTTTPKEPASTTTPKEPTPTTTIKSAPTTPKEPAD--TTTKS   347
                |||          :   |:  ||  |::|     :|  |:   |||   :|:|
Db     72  PGPAGPPGGSRDPG--PPGAPGPAGP--PGSRDPGPPGAPG-PAGPPGGSRDPGPPGAPGPA   126

Qy    348  APTTTKEPAPTTTKEPA-PTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTP---KKPA   403
                |:  |  |||   ||   |:  ||   |||  ||  |||  ||  ||   :|::
Db    127  GPPGSRDPGPPGAPGPAGPPGSRDPGPPGAPGPAGPPGGSRDPGPPGAPGPAGPPGSRDPG   186

Qy    404  PTTPKEPAPTTP---KEPTPTTPKEPA--PTTKEPAPTTPKEPA-------------   442
                |||  ||  ||     |:|  ||  ||   |  ||  ||  ||
Db    187  P--PGAPGPAGPPGGSRDPGPPGAPGPAGPPGSRDPGPPGAPGPKGAHGPAGPKGAHG   244

Qy    443  ------PTAPK-KPAPTTP---KEPAPTTPKEPA--PTTTKEPSPTTTPKEPAPTTTKS   488
                      :     |||  |||   ||  |||  ||   |:  ||  |||   :
Db    245  PAGPKGAHGPAGPKGAPGPAGPPGSRDPGPPGAPGPAGPPGSRDPGP--PGAPGP---A   298

Qy    489  APTTTKEPAPTTTKSAPTTPKEPSPTTTKEPAPTTTKEPAPTTP---KKPAPTTPKEPAP   545
                |:  |  ||   |:  |||  ||   |||  ||  |||  |||      |:  |||  ||
Db    299  GPPGSRDPGP-----PGAPGPAGPPGSRDPGP--PGAPGPAGPPGSRDPGP--PGAPGP   348

Qy    546  TTP---KEPAPTTTKEPA-----PTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTT   595
                |||     ||  |||  ||       |||  ||  |:|  ||  ::::     |:||
Db    349  AGPPGGSRDPGPPGAPGPAGPPGSRDPGP----PGAPGPAGPPGSRDPGP   404

Qy    596  PEKPAPT--TPEELAPTTPEEPTPTTPEEPAPTTPKAAA-PNTPKEPAPTTP---KEPAPT   650
                |  ||||   |  :|  ||   ||  ||   |||          |:  |||      |||  |
Db    405  SRDPGPPGAPGPAGPPGSRDPGP-PGAPGPAGPPGSRDPGPP          462

Qy    651  TPKEPA------PTTTPKET-APTTPKGT---------APTTLKEPAPTTPKKP-APKELA   693
                |||  ||       |:|  |   |:  |||          |||   |:  ||:||   |:|:|
Db    463  GAHGPAGPKGAHGPAGPKGAHGPAGPKGAPGPAGPPGSRDPGPP   522

Qy    694  PTTTKEPTSTTSDKPAPTTPKGTA--PTTPKEPAPTTP---KEPAPTTPKGTA--PTTLKEP   748
                |||   |   :: |||  ||      |:  ||  |||      |||  ||      |:  |||
Db    523  PPGSRDPGPPGA--PGPAGPPGSRDPGPPGAPGPAGPPGSRDPGPPGAPGPAGPPGSRDP   580

Qy    749  APTTPKKPAP------KELAPTTTKGPTST--TSDKPAPTTTPKETAPTTPKEPAPTTPKK   800
                ||  ||:|||      |:|||   |||:     :|  |||    |:|  ||  ||  ||:
Db    581  GP--PGAPGPAGPPGGSRDPGPPGAPGPAGPPGSRDPGPPGAPGPAGPPGSRDPGP--PGA   636

Qy    801  PAPTTP---ETPPPTTSEVSTPPTTTKEPTTIHKSPDESTPELSAEPTPKALEN-----S   851
                |||        |:  |||   |:  |||   :  |    |:|  |      ||
Db    637  PGPAGPPGGSRDPGPPGAPGPAGPPGSRDPGP-PGAPGPAGPPGSRDPGPPGAPGPPG   695

Qy    852  PKEPGVPTTKTPAATK   867
                |  ||  |:
Db    696  SRDPGPPGAHGPAGPK   711
```

```
RESULT 11
US-09-041-886-23
  Sequence 23, Application US/09041886
  Patent No. 6235872
GENERAL INFORMATION:
  APPLICANT:  Bredesen, Dale E.
  APPLICANT:  Rabizadeh, Sharroz
  TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
  TITLE OF INVENTION:  Polypeptides and Methods of Use
  NUMBER OF SEQUENCES:  72
  CORRESPONDENCE ADDRESS:
  ADDRESSEE:  Campbell & Flores LLP
  STREET:  4370 La Jolla Village Drive, Suite 700
  CITY:  San Diego
  STATE:  California
  COUNTRY:  United States
  ZIP:  92122
COMPUTER READABLE FORM:
  MEDIUM TYPE:  Floppy disk
  COMPUTER:  IBM PC compatible
  OPERATING SYSTEM:  PC-DOS/MS-DOS
  SOFTWARE:  PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER:  US/09/041,886
```

```
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Campbell, Cathryn A.
;   REGISTRATION NUMBER: 31,815
;   REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619) 535-9001
;   TELEFAX: (619) 535-9949
; INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1185 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-041-886-23

Query Match      5.8%;  Score 425.5;  DB 4;  Length 1185;
Best Local Similarity  21.8%;  Pred. No. 2.1e-19;
Matches  240;  Conservative  119;  Mismatches  437;  Indels  303;  Gaps  52;

Qy      85 SQTIKSTTKRSRSKPPNKKTKKVIESEEITEEHSVSENQESSSSSSSSSSSSSTIWRIKSS 144
                      ::::                       :::          ::::::::::: :::
Db       9 SMSMRSGRKKEAPGPR----------------EELRSRGRASPGGVSTSSSDG-------KAEKS 50

Qy     145 KNSAANRELQKKLKVKDNKKNRTKKPPTKPPVVDEAGSGLDNGDEKVTTPDTSTTOHNK 204
                      :::                     :::           :::  :::
Db      51 RQTA-------------KKARVEFASTPK---------VNKQGRS------------EEISESESE 82

Qy     205 VSTSPKITTAKPINPRPSLPPNSDTSKETSLTVNKETTVEKETTTINKQTSTDGKEKTT 264
                         :::              :: ::                 :::
Db      83 ETNAPKKTKTEQELPRPQSPSDLDSLDGRSL--NDDGSSDPRDIQQDNRSTS------P 133

Qy     265 SAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTP--KEPASTTP 322
                                      :::          ::: :::::::::
Db     134 SIYSPGSVENDS--------------DSSSGLSQGPA--RPYHPPPLPPSPSQPPDSTP 176

Qy     323 KEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPT 382
                      :::::::::::::::::::::::::::::::::::::::::::::::::::
Db     177 RQPEASF-----------EPHPSVTPTGYHARMEP--PGAPPHPQLYPGGT 224

Qy     383 --TTKSAPTTPK-----------------EPAPTTPKKPAPTTPKEPAPTTPKEPTPTTPK 424
                      :::::::                 ::: ::::::::::::::::::::::
Db     225 GGVLSGPPMGPGPKGGGAASSVGGPNGGKOHPPPTTPISVSSSGASGAPPTKP---PTTPV 280

Qy     425 EPAPTTKEPAP----------TTPKEPAPTA--PKKPAPTTPKEPAPTTPKEPAPTTTKEP 477
                         ::::            :::::::::::::::::::::::::::::::::::
Db     281 GGGNLPSAPPPANPFHVTPNLPPPPALRPLNNASASP----PGLGAQPLPGHL--PSPYA 334

Qy     478 PKE-------PAPTTTKSAPTTTKEP---APTTTKSAPTTP--KEESPTTKEPAPTTPKEPA 528
                      :::           :::        ::::::::::::::  :
Db     335 MGQGMGGLPPGP---EKGPTLAPSPHSLPPASSSAPAPPMRFPYSSSSSSSAAASSSSSS 391

Qy     529 PTTPKKPAPTTPKEPAPTTPKE-PAPTTKKPAPTAPKEPAPTTPKETAPTTPKKLT--- 584
                      : ::::::::::::::::  :::::::::::::::::::::::  :
Db     392 SSSSASPPFAS---QALPSYPHSPPPPTSL-----SVSNQP---PKYTOPSLPSQAVWSQ 440

Qy     585 --PTTPE---KLAPTTPEKPAPTTPEELAPTTPEEPPTPT---------------- 618
                         :::   :::::            :::
Db     441 GPPPPPPYGRRLANSNAHPGPPPSTGAQSTAHPPVSTHHHHHQQQQQQQQQQQQHHG 500

Qy     619 --TPEEPAP------------TTPKAAAPN---TPKEPAPTTPKEP------APTTPKE 654
                      :::                                 :::
Db     501 NSGGPPPGPGAFPHPLEGGSSHAHPYAMSPSLGSLRPYPGPAHLPPPHSQVSYSQAGPNG 560

Qy     655 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP---TTTKEPTSTTSDKPAPT 711
                      : :::       :::::: :::
Db     561 PPVSSSSSNSSSSTSQGSSYPCS--HPSPSQGPGGAPYPPPVPVTTVTTSSAALSTVIATVAS 618

Qy     712 TPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTLKEPAPTTPKKPAPKELAPTTTKGPTS 771
                      :::      :::      ::      :::: ::    :!::::
Db     619 SPAGYKTASPGPPPYGKRAPSPGAYKTATPPGYKPGSP-----PSFRIGTPPGYRG--- 670
```

```
RESULT 12
US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
;   APPLICANT: CHAMBON, Pierre
;   APPLICANT: KIENY, Marie-Paule
;   APPLICANT: LATHE, Richard
;   APPLICANT: HAREUVENI, Mara
;   TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
;   TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;   NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;   STREET: P.O. Box 1404
;   CITY: Alexandria
;   STATE: Virginia
;   COUNTRY: United States
;   ZIP: 22313-1404
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/479,537A
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: FR 90/13101
;   FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO PCT/FR91/00835
;   FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/039,320
;   FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/403,576
;   FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
;   NAME: Teskin, Robin L.
;   REGISTRATION NUMBER: 35,030
;   REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 836-6620
;   TELEFAX: (703) 836-2021
```

```
Qy     772 TTSDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKS 831
                      :  ::: :::           :::::::  ::::     ::::::
Db     671 -TSPPAGPGTFKPGSPIV--GPGPLPPAGPS-GLPSLPPPAAPASGPPLS--ATQIKQE 724

Qy     832 PDE--STPELSAEPTPKALENS--PKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTP 887
                      :::          :::                  :::::
Db     725 PAEEYETPE---SPVPPARSPSPPPKVVDVPSHASQSARFNKHLDRGFNSCARSDLYFVP 781

Qy     888 ETTTAAPKMTKETATTTEKT-TESKLTATTTQVISTTTQDTTPFKIITLKTTTLLAPKVTT 946
                      :::                   ::          ::::
Db     782 ---LEGSKLAKKRADIVEKVRREAEQRAR---------------------------- 807

Qy     947 TKKTTTTTEIMNKPEETAKPKDRATINSKATTPKPQKPTKAPKKPTSTKKPKTMPRVRPK 1006
                      :::::::::::::::::::::::::::::::::::::
Db     808 ----------------EEKEREREREREKEREKEREELERSVKLAQEGRAP-----VECPS 848

Qy    1007 TTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHM 1066
                      :::    ::::      :::: :::
Db     849 LGPVPHR-----PPFEPGSAVA-----TVPPYLGPDTPALRT--LSEYA---------- 885

Qy    1067 LLRPHVFMPEVTPDMDYLP 1085
                      : ::
Db     886 --RPHVMSPGNRNHPFYVP 902
```

```
INFORMATION FOR SEQ ID NO: 5:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1867 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: peptide
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 128..1727
    OTHER INFORMATION: /note= "The amino acids spanning a repeated region wherein the repeat
    OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat
    OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
    OTHER INFORMATION: repeats varies from 1 to 40."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 134
    OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
    OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
    OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 144
    OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
    OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
    OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 147
    OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
    OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
    OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 1..21
    OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
    OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-5

Query Match          5.8%;  Score 424.5;  DB 2;  Length 1867;
Best Local Similarity 23.7%;  Pred. No. 4.2e-19;
Matches  255;  Conservative  69;  Mismatches  511;  Indels  241;  Gaps  53;
```

```
Qy   233 TSLTVNKETTVETKETTTTNKQTTSTDGKEKTTSAKETQSIEKTSAKD-LAPTSKVLAKPT 291
             | |||           |  | ||| ||    |  |  || |   |   | |
Db    16 TVLTV------------------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS 65

Qy   292 P-KAETTTPKGP--ALTTPKEP------------------TPTTTPKEPASTTPKEPTTTIKSAP 334
          | |  ||  ||   | ||| ||                 | ||| ||||| || |||| |
Db    66 PGSGSSSTTQGQDVTLAPATEPASGSAANTWGQDVTSVPVTRPALGSTTPPAHDVT---SAP 122

Qy   335 TTPKEPAPTTT-------KSAPTTPKEP-------------------APTTTKEPAPTTPKEPAPTT 375
          | ||| ||           | ||| ||                  || | | |||| || |
Db   123 --DNKPAPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 180

Qy   376 TKE---PAPTTT--------------KSAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPTPTT 422
          ||   |   |               | ||| |||| || |    | ||| |||| || ||| |
Db   181 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 240

Qy   423 PKE--PAPTTKEP-----------APTTPKEPAPTAPK---KPAPTTPKEPAPTTPKEPAPTT 469
          ||   | || ||           || ||| |||| |     | ||| |||| |||| |
Db   241 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 300

Qy   470 TKEPSPTTPKEPAPTT----TKSAPTTPKEPAPTT---KPAPTTPKEPAPTTPKEPTPTTK----E 518
          | | | ||| |||       || ||| |||| |     | ||| |||| |||| ||| |    |
Db   301 APDXRP-XPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT 359

Qy   519 PAPTTPKEPAPTTPK----KPAPTTKEPAPTTPK----EPAPTTTKKPAPTTAPKEPAPT 570
          | ||| |||| ||        | |||| |||         || |||  || | | |||
Db   360 SAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT 419

Qy   571 TPKETAP----TTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPT 626
             | |    ||   |    | |    |       |   ||    ||     || |
Db   420 SAPDXRPXPGSTAPXAHGVTS----APDXRPXPGSTAPXAHGVTSAPDXRPXPGS 471

Qy    ?  APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRP-XPGSTAPXAHGVTSAPDXRPXPGS 530
Db   472 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRP-XPGSTAPXAHGVTSAPDXRPXPGS 530

Qy   682 TTTP------KKPAPKELAPTTTKEPTSTTSDKPAP--TTPKGTAPTTPKEPAPTTP 729
Db   531 TAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDXRPXPGSTAPXAHGVTS---APDXR 585

Qy   730 KEPAPTTPKG----TAPTTLKEPAPTTP----KKPAPKELAPTTTKGPTSTTSD 775
Db   586 PXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS--TAPXAHGVTSAP 644

Qy   820 TTTTKEPTT------IHKSPDESTPELSAEPTPKALENSPKEPGVPTTTKTPAA---- 865
Db   645 RPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS--TAPXAHGVTSAP 702

Qy   776 KPAP--TTPK---ETAPTTPTPKEPAPTTP---------KKPAPTTPETPPPTTSEVSTP 819
Db   703 DXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAP 762

Qy   866 -TKPEMTTTAKDK---TTERDLRTTTPETTTAAPKMTKETATTTEKTTTESKITATTTQVTS 921
Db   763 DXRPXPGSTAPXAHGVTSAPDXRPXPGST--APXAHGVTSAPDXRPXPGSTAPXAHGVTS 820

Qy   922 TTTQDTTPFKIITTLKTTTLAPKVTTTTKKTTITTTTEIMNKPEETAKPKDRATTNSKATTPKPQ 981
Db   821 APDXRPXPG----STAPXAHGVTI-----SAPDXRPXPGSTAPXAHGVTSAPDXRPXP- 868

Qy   982 KPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT 1041
Db   869 -GSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSA-PDXRPX----------PGST 911

Qy  1042 -PNSKLVEVNPKSEDAGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD 1100
Db   912 APXAHGVTSAPDXRPXPGSTAPXAH---------GVTSAPDXRPXPGSTA----PXAHG 957

Qy  1101 ETNICNGKFVDGLITI-LRNGTLVAFRGHYFWMLSPFSEPSPARRITEVWGIPSPID 1155
Db   958 VTSAPDXRPXPGSTAPXAHGVTSA---------------PDXRPXPGSTAPXAHGVTSAPD 1003
```

```
RESULT 13
US-09-083-116-5
  Sequence 5, Application US/09083116
  Patent No. 6203795
GENERAL INFORMATION:
  APPLICANT: CHAMBON, Pierre
  APPLICANT: KIENY, Marie-Paule
  APPLICANT: LATHE, Richard
  APPLICANT: HAREUVENI, Mara
  TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
  TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
  NUMBER OF SEQUENCES: 5
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
    STREET: P.O. Box 1404
    CITY: Alexandria
    STATE: Virginia
    COUNTRY: United States
    ZIP: 22313-1404
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/083,116
    FILING DATE:
    CLASSIFICATION:
  PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: 08/479,537
  FILING DATE:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 23-OCT-1991
  FILING DATE:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: WO PCT/FR91/00835
  FILING DATE:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/039,320
  FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/403,576
  FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
  NAME: Teskin, Robin L.
  REGISTRATION NUMBER: 35,030
  REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 836-6620
  TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1867 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: peptide
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 134
    OTHER INFORMATION: 128..1727
    OTHER INFORMATION: /note= "The amino acids spanning
    OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat 1
    OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
    OTHER INFORMATION: repeats varies from 1 to 40."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 134
    OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
    OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = Xaa
    OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 144
    OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
    OTHER INFORMATION: which is the codon for Thr or Asn wherein Pro = Xaa
    OTHER INFORMATION: or CCG; and Asn = AAT or AAC."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 147
    OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
    OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
    OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 1..21
    OTHER INFORMATION:
    OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
US-09-083-116-5   21 amino acid precursor sequence."

Query Match
Best Local Similarity  23.7%;   Score 424.5;   DB 4;   Length 1867;
Matches 255;  Conservative 69;  Pred. No. 4.2e-19;  Indels 241;  Gaps 53;

Qy   233  TSLIVNKETTVETKETTTTPNKQTSTDGKKEKTTSAKETQGSIEKTISAKD-LAPTSKVLAKPT  291
               : ||:|                               |:|:|        :| |||
Db    16  TVLTV------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS        65

Qy   292  P-KAETTTPKGP---ALTTPKEP----------TPTTPKEPASTTPKEPTPTTIKSAP       334
            : |:|                             :|:| :|:| ::| :
Db    66  PGSGSSTTGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVT--SAP       122

Qy   335  TTPKEPAPTTT------KSAPTTPKEP----------APTTTKEPAPTTPKEPAPTT        375
            |||:| | :|      ||| :| | |:|          :|  |:
Db   123  --DNKPAPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTS      180

Qy   376  TKE--PAPTTT------KSAPTTPKEPAPTTPK----KPAPTTPKEPAPTTPKEPTPTT     422
               |:| | :|      ||| :|  |:           |:|                 ::|
Db   181  APDXRPXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTS     240

Qy   423  PKE--PAPTTKEP-------APTTPKEPAPTAPK----KPAPTTPKEPAPTTPKEPAPTT    469
               |:|                |:|  |:            |:|                :|
Db   241  APDXRPXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTS     300

Qy   470  TKEEPSPTTPKEPAPTT-TKSAPTTTKEPAPTT------TKSAPTTPKEPSPTTTK----E  518
               |:|  :| | :|    :| | |:|             :| | |:              :
Db   301  APDXRP-XPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVT     359

Qy   519  PAPTTKEPAPTTPK-----KPAPTTPKEPAPTTPK-----EPAPTTTKKPAPTAPKEPAPT  570
            :| |  |:| | :|     |:|         :|:|      :| |                :|
Db   360  SAPDXRPXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVT     419

Qy   571  TPKETAP----TTPKKLTPTTPEKLAPTTPKEPAPTTPEELAPTTPKEPAPTTPEEPAPT    626
            :| |:|     :| | | |: | :|  |:     :| |   :| | |:     :|  :|
Db   420  SAPDXRXPGSTAPXAHGVTS----APDXRPXPGSTAPXAHGVTSAPDXRXPGS         471

Qy   627  TPKA-----AAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGT-APTTLKEPAP          681
            :| |      :| | |:| | :|  |: | :|:|  |:   | |:
Db   472  APXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXR-XPGSTAPXAHGVTSAPDXRXPGS     530

Qy   682  TTP--------KKPAPKELAPTTTKEPTSTTSDKPAP--TTPKGTAPTTPKEPAPTTP     729
            :|         |:| |  :|  |:|      :| |    :| |: | :| | :|  :|
Db   531  TAPXAHGVTSAPDXRXPGSTAP-XAHGVTSAPDXRXPGSTAPXAHGVTS----APDXR     585

Qy   730  KEPAPTTPKG---TAPTTLKEPAPTTP-----------KKPAPKELAPTTTKGPTSTTSD    775
              |:| | :|    |:|  |:| | :|            |:| |  :|  |:|
Db   586  PXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDX      644

Qy   776  KPAP---TTPK----ETAPTTPKEPAPTTP--------KKPAPTTPETPPPTTSEVSTP     819
            |:|     :| |    :|  |: | :| | :|        |:| |:|   :| |
Db   645  RPXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXRXPGS--TAPXAHGVTSAP     702

Qy   820  TTTKEPTT-------IHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA-----    865
            :|  |:|        :| |    :| |   |: |    |:|      | :| |
Db   703  DXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVTS      762

Qy   866  -TKPEMTTTPAKDK---TTERDLRKTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTS  921
               |  |: | |:|    :| |  |:| |: |  |:|  |:|  |: |:|  |:  |: | :|:
Db   763  DXRXPGSTAPXAHGVTSAPDXRXPGSTAPXAHGVT-----SAPDXRXPGSTAPXAHGVTSAP  820

Qy   922  TTTQDTTTPFKITTLKTTTLAPKVTTWTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQ  981
            :|    :| |  :| | :|  |:|       :|  |  |:|      |:|      |:   |:|
Db   821  APDXRPXPG------STAPXAHGVT------SAPDXRXPGSTAPXAHGVTSAPDXRXP-     868

Qy   982  KPTKAPKKPTSTKKPKITMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT  1041
            |:|      |:|      |: |       |:|  |:|         | |      :|
Db   869  -GSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSA-PDXRPX------------PGST    911

Qy  1042  -PNSKLIVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD  1100
            :| |       |:|      |:|      |: |       |:|       | |
Db   912  APXAHGVTSAPDXRPXPGSTAPXAH--------GVTSAPDXRPXPGSTA-----PXAHG    957

Qy  1101  ETNICNGKPVDGLTT-LRNGTLVAFRGHYFWMLSPSPSPARRITEVWGIPSPID        1155
            :|      |: |  :| |  |: |       |:|             | |
Db   958  VTSAPDXRPXPGSTAPXAHGVTSA----------PDXRPXPGSTAPXAHGVTSAPD      1003
```

RESULT 14
US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Patent No. 5861381
GENERAL INFORMATION:
  APPLICANT: CHAMBON, Pierre
  APPLICANT: KIENY, Marie-Paule
  APPLICANT: LATHE, Richard
  APPLICANT: HAREUVENI, Mara
  TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
  TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1899 constitute a repeated region wherein the repeat i
CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
20 amino acids, 17 of which are fixed. The number of such
repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Qy   Query Match
      Best Local Similarity   5.8%;   Score 424.5; DB 2;   Length 2035;
Matches 255; Conservative 69; Mismatches 511;   Indels 241; Gaps 53;
                                   23.7%; Pred. No. 4.6e-19;

Qy   233 TSLTTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTTSAKD-LAPTSKVLAKPT 291
           :||| : ||| : :| |:||| |||:|:::|:::|::::| :|| :|: :
Db   16   TVLTV-------------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS 65

Qy   292 P-KAETTTKGP---ALTTPKEP-----------------TPTTTPKEPASTTPKEPTPTTIKSAP 334
           : :||:| || | | |||||:||| ||||:| |||: :| :|: ||
Db   66   PGSGSSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVT---SAP 122

Qy   335 TTPKEPAPTTT--------KSAPTTPKEP---------------APTTTPKEPAPTTPKEPAPTT 375
           | : | :|: : |: :|| |:
Db   123 --DNKPAPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 180

Qy   376 TKE--PAPTTT------KSAPTTPKEPAPTTPK----KPAPTTPKEPAPTTPKEPTPTT 422
           ||| |||| ||| ||| ||| ||| |:
Db   181 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 240

Qy   423 PKE--PAPTTKEP------APTTPKEPAPTAPK----KPAPTTPKEPAPTTPKEPAPTT 469
           ||| | ||:: ||| ||| ||| | |:: |||
Db   241 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS 300

Qy   470 TKEPSPTTPKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTPKEPSPTTTK----E 518
           ||:| | |:   :||| | |: ||| |: |: |
Db   301 APDXRP-XPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT 359

Qy   519 PAPTTPKEPAPTTPK-----KPAPTTPKEPAPTTPK----EPAPTTKKPAPTTAPKEPAPT 570
           | || || ||| ||| ||| | |||: | |||
Db   360 SAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT 419

Qy   571 TPKETAP-----TTPKKLTPTTPKEPAPTTPEELAPTTPEEPTPTTPEEPAPT 626
           | ||: ||| | |: |: ||| | |: |:
Db   420 SAPDXRPXPGSTAPXAHGVTS----APDXRPXPGSTAPXAHGVTSAPDXRPX-----PGST 471

Qy   627 TPKA-------AAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGT-APTTLKEPAP 681
           | || | |: |: | | |: |: || | |||
Db   472 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSAPDXRPXPGS 530

Qy   682 TTP-----------KKPAPKELAPTTTKEPTSTTSDKPAP--TTPKGTAPTTTPKEPAPTTP 729
           | | | |||:| |:| |: ||| ||| || | |||
Db   531 TAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDXRPXPGSTAPXAHGVTS-----APDXR 585

Qy   730 KEPAPTTPKG----TAPTTLKEPAPTTP----------------KKPAPKELAPTTTKGPTSTTSD 775
           |:|| | ||| | |||| | |||:| |:|
Db   586 PXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS----APDXRPXPGSTAP-XAHGVTSAPDX 644

Qy   776 KPAP----TTPK---ETAPTTPKEPAPTTP-------------KKPAPTTPETPPPTTSEVSTP 819
           | | ||| | ||| |: | | ||| | | |||: :|
Db   645 RPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS--TAPXAHGVTSAP 702

Qy   820 TTTTKEPTT-----------IHKSPDESTPELSAEPTTPKALENSRKEPGVPTTKTPAA---- 865
           || ||:| |:| | | ||| | | ||
Db   703 DXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAP 762

Qy   866 -TKPEMTTTAKDK---TTERDLRTTPETTTAAPKMTKETATTTEKTTTESKITTATTTQVTS 921
           |:| || | |:| |||| | |:|: |:| | |: | | |:
Db   763 DXRPXPGSTAPXAHGVTSAPDXRPXPGST-APXAHGVTSAPDXRPXPGSTAPXAHGVTS 820

Qy   922 TTTQDTTPFKITTLKTTTLAPKVTTTKKTTITTTEIMNKPEETTAKPDRATNSKATTPKPQ 981
           | | || :|: |||| | :|||:| |::
Db   821 APDXRPXPG-----STAPXAHGVT------SAPDXRPXPGSTAPXAHGVTSAPDXRPXP- 868

Qy   982 KPTTKAPRKPTTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT 1041
           |||:| |:| : | |: :||: ||| | |:| || ::| |: ||
Db   869 -GSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSA-PDXRPX-----------PGST 911

Qy   1042 -PNSKLVEVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYLPRVPNQGIIINPMLSD 1100
           |: || |:

```
Db    912 APXAHGVTSAPDXRPXPGSTAPXAH--------GVTSAPDXRPXPGSTA-----PXAHG  957
Qy   1101 ETNICNGKPVDGLTT-LRNGTLVAFRGHYFWMLSPFSPSPSPARRITEVWGIPSPID   1155
Db    958 VTSAPDXRPXPGSTAPXAHGVTSA--------PDXRPXPGSTAPXAHGVTSAPD     1003
```

RESULT 15
US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795

GENERAL INFORMATION:
APPLICANT: CHAMBON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREUVENI, Mara
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
  ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
  STREET: P.O. Box 1404
  CITY: Alexandria
  STATE: Virginia
  COUNTRY: United States
  ZIP: 22313-1404
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/083,116
  FILING DATE:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/403,576
  FILING DATE: 14-MAR-1995
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/039,320
  FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: WO PCT/FR91/00835
  FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/479,537
  FILING DATE:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER:
  FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
  NAME: Teskin, Robin L.
  REGISTRATION NUMBER: 35,030
  REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 836-6620
  TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 2035 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: peptide
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 128..1899
    OTHER INFORMATION: /note= "The amino acids spanning
      128 to 1899 constitute a repeated region wherein the repeat is
      20 amino acids, 17 of which are fixed. The number of such
      repeats varies from 1 to 40."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 134
    OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
      Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
    OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 144
    OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
      which is the codon for Thr or Asn wherein Thr = ACT, ACC,
    OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 147
    OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
      which is the codon for Pro or Ala wherein Pro = CCT, CCC,
    OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
  FEATURE:
    NAME/KEY: Peptide
    LOCATION: 1..21
    OTHER INFORMATION: /note= "amino acid precursor sequence."
US-09-083-116-2
    21 amino acids 1 to 21 are a
```

```
Query Match
Best Local Similarity   23.7%;   Score 424.5; DB 4;   Length 2035;
Matches 255;   Conservative 69;   Mismatches 511;   Indels 241;   Gaps 53;
```

```
Qy  233 TSLTVNKETTVETKETTTNKQTSTDGKEKTTSAKETQSIEKTSAKD-LAPTSKVLAKPT  291
         |||                   ||| |||  ||| |||  || ||   |  ||
Db   16 TVLTV--------------VTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS  65

Qy  292 P-KAETTTKGP--ALTTPKEP----------------TPTTPKEPASTTPKEPTPTTIKSAP  334
         | : ||  |:    |  |||                 ||  |
Db   66 PGSGSSTTQGGDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSSTTPPAHDVT---SAP  122

Qy  335 TTPKEPAPTTT----------KSAPTTPKEP--------APTTKEPAPTTPKEPAPTT  375
         :||:.:|              |  ||               ||  ||     |
Db  123 --DNKPAPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  180

Qy  376 TKE--PAPTTT---------KSAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPTPTT  422
         |:    |              |||     ||  |      |
Db  181 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  240

Qy  423 PKE--PAPTTKEP--------APTTPKEPAPTAPK---KPAPTTPKEPAPTTPKEPAPTT  469
         |:    ||             ||     |          |
Db  241 APDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTS  300

Qy  470 TKEPSPTTPKEPAPTT--TKSAPTTKEPAPTT----TKSAPTTPKEPSPTTTK----E  518
         |    |             |||           |              |||
Db  301 APDXRP-XPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT  359

Qy  519 PAPTTPKEPAPTTPK----KPAPTTPKEPAPTTPK----EPAPTTIKKPAPTAPKEPAPT  570
         |   |                |                  |
Db  360 SAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVT  419

Qy  571 TPKETAP-----TTPKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPT  626
         |   |       |: |                 |
Db  420 SAPDXRPXPGSTAPXAHGVTS----APDXRPXPGSTAPXAHGVTSAPDXRPX----PGST  471

Qy  627 TPKA----AAPNTPKEPAPTTPKEPAPTTPKEPAPTTPKGT-APTTLKEPAP  681
         | :     |:|                               |
Db  472 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRP-XPGSTAPXAHGVTSAPDXRPXPGS  530

Qy  682 TTP----------KKPAPKELAPTTKEPTSTTSDKPAP--TTPKGTAPTTPKEPAPTTP  729
         | |           |   |  |                 ||
Db  531 TAPXAHGVTSAPDXRPXPGSTAP-XAHGVTSAPDXRPXPGSTAPXAHGVTS----APDXR  585

Qy  730 KEPAPTTPKG----TAPTTLKEPAPTTP----------KKPAPKELAPTTTKGPTSTTSD  775
         |   |          | |                     |   |  |
Db  586 PXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDX  644

Qy  776 KPAP--TTPK---ETAPTTPKEPAPTTP----------KKPAPTTPETPPPTSEVSTP  819
         | |    ||                               |   | |
Db  645 RPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS--TAPXAHGVTSAP  702
```

```
Qy   820 TTTKEPTT--------IHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAA------- 865
             :|    |    |  |  : :|| | |: :|| | |
Db   703 DXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAP 762

Qy   866 -TKPEMTTTAKDK---TTERDLRTTPETTTAAPKMTKETATTTEKTTESKIITATTTQVTS 921
              :|| ||      | ||       | |    |||
Db   763 DXRPXPGSTAPXAHGVTSAPDXRPXPGST--APXAHGVTSAPDXRPXPGSTAPXAHGVTS 820

Qy   922 TTTQDTTPFKITTLKTTTLAPKVTTTKKTTITTTEIMNKPEETAKPKDRATNSKATTPKPQ 981
             :    :          |||
Db   821 APDXRPXPG-----STAPXAHGVT-----SAPDXRPXPGSTAPXAHGVTSAPDXRPXP- 868

Qy   982 KPTTKAPKKPPTSTKKPKTMPRVRKPKTTPTPRKMTISTMPELNPTISRIAEAMLQTTTRPNQT 1041
             :| |      |:  |:            |
Db   869 -GSTAPXAHGVTSAPDXRP--XPGSTAPXAHGVTSA-PDXRPX----------PGST 911

Qy  1042 -PNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSD 1100
              | |         :            ||         :
Db   912 APXAHGVTSAPDXRPXPGSTAPXAH--------GVTSAPDXRPXPGSTA----PXAHG 957

Qy  1101 ETNICNGKPVDGLTT-LRNGTLVAFRGHYFWMLSPFSPPSPARITEVWGIPSPID 1155
             | |             |  |           |:  |   |
Db   958 VTSAPDXRPXPGSTAPXAHGVTSA----------PDXRPXPGSTAPXAHGVTSAPD 1003
```

Search completed: April 26, 2002, 16:25:33
Job time: 473 sec

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: April 26, 2002, 16:31:27 ; Search time 114.61 Seconds
                 844.095 Million cell updates/sec
                 (without alignments)

Title: AA3
Perfect score: 6814
Sequence: 1 MAWKTLPIYLLLLLSVFVIQ..........ARAITTRSGQTLSKVWYNCP 1270

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_68:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	1065	15.6	3020	2	A43932	mucin 2 precursor,
2	950	13.9	1664	2	T18262	S-layer protein,
3	855	12.5	1489	2	T31108	cyst germination s
4	802	11.8	1274	2	T16251	hypothetical prote
5	798.5	11.7	1367	1	S48478	glucan 1,4-alpha-g
6	784	11.4	1188	2	S49915	extensin-like prot
7	776.5	11.4	2187	2	T30826	nascent polypeptid
8	672	9.9	1344	1	A35175	mucin 1 precursor,
9	671	9.9	1151	2	T18535	high molecular mas
10	670	9.8	1229	2	T25697	hypothetical prote
11	651.5	9.8	3570	2	T45025	mucin MUC5B, trach
12	641.5	9.4	3507	2	T34513	hypothetical prote
13	635.5	9.3	7962	2	I38346	elastic titin - hu
14	633	9.3	489	2	T11622	extensin class 1 p
15	632	9.3	761	2	C84672	hypothetical prote
16	630.5	9.3	990	2	I51618	nucleolar phosphop
17	626.5	9.2	971	2	T19431	hypothetical prote
18	625	9.2	6642	2	T29757	protein UNC-89 - C
19	622.5	9.1	839	2	F75518	hypothetical prote
20	608	8.9	801	2	T29018	hypothetical prote
21	607.5	8.9	924	2	S27923	gene LF3 protein -
22	605	8.9	379	2	S50125	larval glue protei
23	559.5	8.2	350	2	S22456	hydroxyproline-ric
24	556.5	8.2	856	2	T16543	hydroxyproline-ric
25	556.5	8.2	875	2	S23760	polyphenolic adhes
26	555.5	8.2	1630	2	A53577	ascites sialoglyco
27	554	8.1	2232	2	T34434	hypothetical prote
28	551	8.1	620	2	S06733	hydroxyproline-ric
29	545	8.0	873	2	A47203	calphotin - fruit
30	542	8.0	369	2	S20500	hydroxyproline-ric
31	540.5	7.9	756	2	T27642	hypothetical prote
32	538.5	7.9	416	2	JU0465	extensin precursor
33	533	7.8	1087	2	QFMSH	neurofilament trip
34	530.5	7.8	1162	1	JH0557	exo-alpha-sialidas
35	530	7.8	865	2	A47282	calcium-binding pr
36	522	7.7	328	2	JQ0985	hydroxyproline-ric
37	522	7.7	1459	2	T32271	hydroxyproline-ric
38	518	7.6	866	1	T45462	hypothetical prote
39	518	7.6	1072	2	S70795	membrane glycoprot
40	512.5	7.6	867	2	A37221	neurofilament trip
41	509.5	7.5	662	1	T45463	membrane glycoprot
42	509.5	7.5	1832	2	A45155	mucin FIM-C.1 - Af
43	508	7.5	505	2	T31113	hypothetical prote
44	505	7.4	700	2	A54641	mucin-like glycopr
45	504.5	7.4	606	2	A43427	neurofilament trip

ALIGNMENTS

RESULT   1
A43932
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
A:Accession: A49963; A45106; A43932; B33532; A61257; PQ0328; PQ0329
J:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J:Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MUID:94132002
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GUI>
A:Cross-references: GB:L21998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J:Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up |
A:Reference number: A45106; MUID:93016075
A:Accession: A45106
A:Molecule type: not compared with conceptual translation
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J:Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:91358717
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L',1352-1411, 'S',1413-1448, 'P',1450-1503, 'T',1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J:Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A:Reference number: A33532; MUID:89197956
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

```
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481
A;Accession: A61257
A;Molecule type: mRNA
A;Status: not compared with conceptual translation
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A;Reference number: PQ0328; MUID:92198477
A;Accession: PQ0328; MUID:92198477
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C;Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>


Query Match    15.6%;   Score 1065;  DB 2;   Length 3020;
Best Local Similarity   28.7%;   Pred. No. 2e-39;
Matches  370;   Conservative 79;   Mismatches 464;   Indels 376;   Gaps 49;

Qy    29 CKGRCFE-----------------------------SFERGRBCDDAQCKKYDKCCPDYESFCAEVKDNKKNRI 74
            | ||                             : | ||
Db  1137 CGNRSFETCRTINGIHSNISVSVLEG----CYPRCPK-DR-PIYE-----EDLKKCVI 1183

Qy    75 KKK------PTPKPP---------------VV--DEAGSGLD--NGDF----KVT 100
            ||          || ||            : ||            ::
Db  1184 ADKGCGCYVEDTHYPGASVPTEETCKSCVCTNSSQVVCRPEEGKILNQTQDGAFCYWEIC 1243

Qy   101 TPDFSTTQH---NKVSTSPKITVTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKET-- 155
            :|||          :|| ||           :|||
Db  1244 GPNGTVEKHFNICSITTRPSTLTTFTTITLPTTP------TSFTTTTTTTTPISSTVL 1295

Qy   156 TTTNK-----------------QTSTDGKEKTTS--AKETQSIEKTSAKD---- 186
            ::||                                      ||       ||
Db  1296 STTPKLCCLWSDWINEDHPSSGSDGDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKV 1355

Qy   187 --------------------LAPTSKVLAKPTPKAETTTK 206
                                :|:|      || |||
Db  1356 QCDVSVGFICKNEDQFGNGPFGLCYDYKIRVNCCWPMDKCCITTPSPPTTTTSPPTTTTT 1415

Qy   207 GPALTTPKEPTPTTPKEPASTTPKEPTPTTPTIKSAP--TTPKEPAPTTTKSAPTTPKEPAPT 265
            | :|      | |||                 : :|| |||
Db  1416 LPPTTPSPPTTTPTTTTTTTPSPPTTTTTTTTPPTTTTPLPTTTPSPPISTTTTPPTTTTPSP-PT 1474

Qy   266 TTKEPAPTTPKEPAPTTTKEPAPTTTKSAP--TTPKEPAPTTTKKPAPTTTPKEPAPTTPKE 324
            ||      |||           : ||| |||         : |||
Db  1475 TTPSPPTTPSPPT-TTTTTPPTTTTTTTPSPPMTTPTTTPPASTTLPPTTTPSPPTTTP 1533

Qy   325 PTPTTPKEPAPTTKEPAPTTAPKKPAPTTTPKKETAPTTPKKLTPTTPPEKLAPTT 384
            | |||       || |||                        :|
Db  1534 PPTTTTPSPPTTTPITP-PTSTTTLPPTTTPSPPTTTTPSPPTTTPSPPTTT 1592

Qy   385 PKEPAPTTTIKSAP--TTTPKEPAPTTTIKSAP--TTPKEPSPTTTTKEPAPTTTPKKP 442
            | | |||  : ||  |||        : ||   |||           ||
Db  1593 TTTTPPPPTTTPSPPTTTTTTTTTPPTTTPIPPTTTPSPPTTTPSPPTTTTP 1652

Qy   443 APTTTPKEPAPTTPKKPAPTTPKKPAPTTPKEFAPTTPKKLTPTTPKEKLAPTT 502
            |||      | |||          || ||
Db  1653 PTTTPSPPTTTTPSPITTTTTPPTTTPPSPITTTTPPTTTTPSS-PITTT--TMTTPS--PTTT 1704


RESULT    2

T18262
S-layer protein - Clostridium thermocellum
C;Species: Clostridium thermocellum
C;Date: 15-Oct-1999    #sequence_revision 15-Oct-1999    #text_change 15-Oct-1999
C;Accession: T18262
R;Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A;Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulo
e.
A;Reference number: Z18847; MUID:93209931
A;Accession: T18262
A;Status: Preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1664 <FUJ>
A;Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1


Query Match    13.9%;   Score 950;   DB 2;   Length 1664;
Best Local Similarity   31.0%;   Pred. No. 1.2e-34;
Matches  312;   Conservative 104;   Mismatches 361;   Indels 230;   Gaps 55;

Qy   193 VLAKPPP-KAETTTKGPALTTPKEPPTTPKEPASTTPKEPAPTTIKSAPTTPKEPAPTT 251
            ||  |   ||          |||||                     : ||
```

```
Qy   503 PEKPAPTTPEELAPTTPEEPPTTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEP 562
            | |||     | |||                     |||      |||
Db  1705 PSSPITTTTPPTTTPSSTTTP-SPPTTTMTTPSPPTTTTPSPPTTTMTTLPPTTTTSSPLTTTPLPP 1763

Qy   563 APTTTP-----KETAPTTP-------------------KG------ 577
            |||       ||                          |
Db  1764 SITPPTTSPFSTTTPTTTPCVPLCNWTTGWLDSGKPNFHKPGGDTELIGDVCGPGWAANISC 1823

Qy   578 -----------------------TAPTT 581
                                   :| ||
Db  1824 RATMYPDVPIGQLGQTVVCDSVGLICKNEDQKPGGVIPMAFCLNYEINVQCCECVTQPT 1883

Qy   582 TLKEPAPTTPKKPAPKELAPTT--TKEPTSTTYSDKP----APT-TPKGTAPTTPKEPAPT 634
            | |||       | ||       :|             |||      |||
Db  1884 TM---TTTTTTENPTTPITTTTTTVTPTPPTTPTPTPTSIQSPNGLQAPTTPISTTTTTVTPTPTPTT 1940

Qy   635 TPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTST-TSDKPAPTTPKETA 693
            | |||       || |||       | |||                      |||
Db  1941 GTQTPT-TPIFITTTTTTVTPTPTPTGTQTPTTVLITTTTTTMTPTPTPTSTKSTTTVTPITTT 1999

Qy   694 PTTTPKEPAPTTPKKP---------APTTPETPTPPPTTSEVSTPTTTK-----EPTTIHKSPDE 741
            | |||                         | ||              ::
Db  2000 TTVIATPTTPTGTQTPTTMIFISTTTTVTPTPTPTTPTTGSTGPPTTHTSTAPTIAELTTTSNPPPES 2059

Qy   742 STPELSAEPTPKALENSPKEPGVPTTKTPAANTKPEMTTTAKDKTTERDLRTTPPETTTAAP 801
            | |      |                   :||            :|
Db  2060 STPQTSRSTSSPLTESTTLLSTLPPAIEMTSTAPPSTPTAPTTSGGHTLSPPPSTTTTSP 2119

Qy   802 KMTKETATTTTEKTTPESKITATTTQVTSTTTQDTTPFKITT----LKTTTLAPKVTTTKKTI 858
            | | |               :|                       :| |||
Db  2120 PGTPTPRGTTT--GSSSAPPTSTVQTTTTSAWTPTPTPTLSTPSIIRTTGLRPYPPSSVLICC 2177

Qy   859 TTTEIMNKPEE------------------TAKPKDRATNSKATTPKQK 889
            ||                          |||
Db  2178 VLNDTYYAPGEEVYNGTYGDTCYFVNCSLSCTLEFYNWSCCPSTPSPTTPPSK-STPTPSK 2236

Qy   890 PTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTP 949
            | |          :||                              |
Db  2237 PSSTPSKPTPGTKPPECPDPDPR------------------QENETWWLCDCFM-ATCKYNNTV 2282

Qy   950 NSKLVEVNP------------KSEDAGG-------------AEGETPHMLL-------- 975
            ::|                  |                   :|
Db  2283 EIVKVECEGPPPMPTCSNGLQPVREDEPDGCCWHWECDCYCTGWGD-PHYVTFDGLYYSYQ 2341

Qy   976 --RPHVFMPEVTPDMDYLPRVPNQGIIIN 1002
            ||              :|:
Db  2342 GNCTYLVLVEEISPSVD----NFGYVID 2364
```

Query Match   12.5%;   Score 855;   DB 2;   Length 1489;
Best Local Similarity   31.8%;   Pred. No. 1.6e-30;
Matches 362;   Conservative 52;   Mismatches 496;   Indels 228;   Gaps 51;

```
Db    758  VVIQPAPIKAASDEPIPTDPSDEPTPS----------DEPTPS---DEPTPSDEPTPSD  804
                ||:                                           ||::|
Qy    252  TKSAPTTPKEPAPTTTKEPAPTTPKEPA---------PTTTKEPAPTTPKSAPTTPKEPAPTTPK  307

Db    805  EPTPSETPEEPIPTDTPSDEPTPSDEPTPS  864
           ||:
Qy    308  KPAPTTPKEPAPTTPKEPT-TPKEPAPTAPKKPAPTTPKEPAP  366

Db    865  SDEPTPS--DEPTPS--DEPTP-SDEPTPSETPEEPIPTDTPSDEPTPSDEPTPP  919
              ||||         |:||       ::|
Qy    367  TTPKEPA--PTTTKEPSPT-TPKEPAPTTTKS-APTTTKEPAPT---TTKSAPTTPKEPS  419

Db    920  SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT  979
              ::                   :|
Qy    420  ----PTTTKEPAPT-TPKEPAPTTPKKPAPTTPKEPAPTTTKKPAPTTAPKEP  474

Db    980  PSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEP  1039
                                 :|
Qy    475  APT--------TPKETAPTTPKKLTPTTPEKLAPTTPEKPAPT-----TPEELAPTTPEEPTPT-TP  527

Db   1040  TPSDEPTPSE--TPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSETP  1094
              ||                     :|:
Qy    528  EEPAPTTPKAAAENTPKEPAPTTPKEPA--PTTPKEPAPT-TPKEPAPTTPKGTAPTTLK  584

Db   1095  EEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSD  1154
                                                          :|
Qy    585  EPAPTTPKKPAPKELAPTTPKEPT-STTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTT  643

Db   1155  EPTPS--DEPTPSD-EPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPS-  1208
           ||||            |||
Qy    644  PKGTAPTTLKEPAPT-TPKKPAPKELAPTTPKEPT---TPKETAPTTPKE  699

Db   1209  ---DEPTPSDEPTPSETPEEPI--  1253
                 ||:|||
Qy    700  PAPTTPKKPAPT-TPEPPPPTTPSEVSTPTTTPFKITILKTTILAPKVTTTKKTITTTEIMNKPEETAKPKDR  758

Db   1254  PTPS--DEPTPSDEPTPSDEPTP--SDEPTPSDEP  1305
              ::
Qy    759  PKEPGVPTTKTPAAT--KPEMTTTAAKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTE  816

Db   1306  SDEPTPSDEPTPSDEPTPSD---EPTPSDEPTPSDEPTPS  1361
                 ::
Qy    817  SKITATTTQVTSTTTQDTTPFKITILKTTILAPKVTTTKKTITTTEIMNKPEETAKPKDR  876

Db   1362  DEPTPSDEPTPSETPEEPTP---TTTPTPTSTT  1396
                                     ||| |:
Qy    877  ATNSKAT-----TPKPQKPTKAP-KKPTSTKKPKTMPRVRKPKTTPTPRKM------  921

Db   1397  SGGSGSGGSGGGGGGGTVPFTSPTPTPTSKPTSTPAP---TEIEEPTPSDVPGAIGGEHRA  1453
                                                      :| ::
Qy    922  ------TSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAEGETPHM------  973

Db   1454  YLRGYPDGSFRPERNITRAEAAVIP-----AKLL---GADESYGAQSASPYSD  1498
                                         :|:|      :
Qy    974  LLRPH--------VFMPE----------VTPDMDYLPRVPNQGI-----  999

Db   1499  LADTHWAANAIKFATSQGLFKGYPDGTFKPDQNITRAEPATVVLHFLTKVKGQEIMSKLA  1558
                  ::                 :   |         ::
Qy   1000  ---IINPMLSDETNICNG----RPVDGLTTL-----RNGTLVAFRGHYFWMLSPFSPPS  1046

Db   1559  TIDISNPKFDD-----CVGHWAQEFIEKLTSLGYISGYPDGT-----  1600
                 ::            ::|:
Qy   1047  PARRITEVWGIPSPIDVVFTRCNCEGKTFFFKD--SQVWRFTNDIKD  1091

Db   1601  YIKRSESV----ALINRALERGPLNGAPKLFPDVNESYWAF-GDIMD  1642
```

```
Qy     37  FERGREGCDCDAQCKKFXDKCCPDYESFCCAEVKDN--KKNRTKKKPTPKPPVVDEACSGLD  93
           |:|                      ::                |:
Db    301  YEPSDETEAPTEGTTY---VPREETTAAPSEDTTYAPAEEKPVDVEETTYVTE  357

Qy     94  NGDFKVTTFPDTSTT-----QHNKVSTSPKITTAKPINPRPSLP----  132
              ||                    ::::
Db    358  ESTYAPPIKSETNAPTERMHYAHIEKPCDTEVTMYAPTEETTYAPIEETTYAP  417

Qy    133  NSDT----SKETSLTVNKETTVETTKETTT---TNKQTSTDGKEKTTSAKETQSIEKTSAK  185
                    ::|:                        :
Db    418  TEETPYEPTEETTYIPTEETTYAPTEETTYAPTEETTYAPTEEKTTYAPTEETTYAPTEET  477

Qy    186  DLAPTSKVLAPKPTPKAETTTKGPALITTPKPTTPKEPASTTPKEPTTPTTIKSAPTTPK  245
                                         :|          |:
Db    478  TYAPIKETTYAPT---EETTYASTEETTYAPTEETTYAPAEETTYEPTEET-TYAPTEET  533

Qy    246  EPAPT-TTKSAPTTPKEPAPT---TTKEPAPTTPKEPAPTTTKEPAPTTTKSAPT-----  296
                   |||               :|
Db    534  TYAPTEETTYAPTEETTYAPAEETPYEPTEETTYAPAEETTYAPAEETTYA-YAPTEETMYA  592

Qy    297  ----TPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPPTTPKEP---AP---  335
               :|
Db    593  PIEETTYAPTEETTYAPAEETPYEPTEETTYAPTEETTYAPTEE  652

Qy    336  TTTKEPAPTTPKEP-----APTAPKKPAPTTPKEPAPTT-TTKEPSPTTPKE  387
                  :|
Db    653  TTYAPAEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA  712

Qy    388  P--------APT-TTKSAPT-----TTKEP-----APT-TTKSAPT-TPKEPSPT  421
                       |:
Db    713  PTEETTYAPTEETTYAPTEETTYAPTEETTTYAPTEETMYAPTEEETTYAPTEE  772

Qy    422  TTKEP------APTTPKEPAPTTPKEPAPT--TPKEPAPTTKKPAPTTPKE  473
                       :|
Db    773  TTYEPTGEETTYAPTEETTYAPIEETTYAPTEEEATTYAPTEEETTYAPTEE  832

Qy    474  PAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPT--TPEEPPPTTPEEP-  530
                                                          :|
Db    833  PTEET--TYTPTEETTYAPTEETKTTYAPTEETTYAPTEEETPYEPTEETTYAPT  890

Qy    531  -----APT--TPKAAANTPKEPAPTTPKEP-----APTTPKEPAPTTPKE  569
                :|
Db    891  KETTYAPTEETTYASTEETTYAPAEETTYAPTEETTYAPIEET  950

Qy    570  TAPTTPKGTAPT--TLKEPAPTTPKKPAPK-ELAPT--TTKEPTSTTSDKPAPTTPKGTA  624
                                           :|
Db    951  YAPTEETTYAPIEETTYAPAEEETPYEPTEETTYAPTEE  1010

Qy    625  PTTPKEPAPTTPKEP-----APTTPKGTAPT-----TLKEPAPTTP  660
                :|
Db   1011  EETTYAPAEETPYEPTEETTYAPTEETTYASTEETTYAPAEETP  1070

Qy    661  KKP-----APKE---LAPT--TTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKPA  709
              |                :|
Db   1071  YEPTEETTYAPTEETTYAPIEETTYAPAEETPYEPTEETTYAPT  1130

Qy    710  PTTPEPPPPTTSEVSTPTTTKEPT--TIHKSPDESTPELSAEPTPKA-LENSPKEPG---  763
```

```
Qy  1131  EETTYAPTEETMYAPIEETTYGPTEETTY-APTEATTYAPTEETPYAPTEETTYEPIGET  1189
          : :|    |:| :                |    |::| :|   |      ||
Db   764  --VPTTKTPAATKPEMTTTAKIDKTTERDLRTTP----ETTTAAPKMTKETATTEKTTES   817

Qy  1190  TYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPIEETTYAP----TEETTYEPTEE   1244
          |   |  |           |:|                            |||| ||
Db   818  KITATTTQVTSTTTQDTT----------PFKITT---LKTTTLAPKVTTTKKITTTEIM    864

Qy  1245  TTYYAPTEETTYAPTEETTYAPTEETMYAPIDETTYGPTEETTYAPTEATTYAPTEETPYA  1304
          | | | ||| |     |   |  | |   |  :|                   | |
Db   865  NKPEETAKPKDRAT--NSKATTPKP-QRPTKAPKKPT------------STKKP-KTMPR    908

Qy  1305  PTEETTYEPIGETTYAPTEETTYAPTEETTYAPMEEETPYEPAEESTSTVSTEKPCNTEEF  1364
          |  |  | :| :|  |     |   :|    |    |  |   :||       |:|| :|
Db   909  VRKPKTTPTPPRKMTSTMPELNPTSRIIAEAMLQTTTRP--NQTPNSKLVEVNPKSEDAG   964

Qy  1365  TDEPTDEPTDE--PSDEPTDEPTDLPTDEPSTPCDNQGINGIGVENKVRTNNAG        1420
          |::| |: | :    |  |        |::| : :|   | :|: |  |:
Db   965  ...
```

RESULT   4
T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Accession: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
R:Leimbach, D.
submitted to the EMBL Data Library, January 1996
A;Description: The sequence of C. elegans cosmid F35A5.
A;Reference number: Z18485
A;Accession: T16251
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <LET>
A;Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GN00
A;Experimental source: strain Bristol N2; clone F35A5
A;Genetics:
A;Gene: CESP::F35A5.1
A;Map position: X
A;Introns: 1272/2

```
Query Match     11.8%;  Score 802;  DB 2;  Length 1274;
Best Local Similarity   28.0%;  Pred. No. 2.9e-28;
Matches  290;  Conservative  102;  Mismatches  414;  Indels  230;  Gaps  56;
```

```
Qy   71  KNRTKK-KP------TP------KPPVVDEAGSGLDNGD----FKVTTPDTSTTQHNKVTS  115
         |: ||    |            |  || |   | |         | |  | |::|:
Db  274  KNPTKKWKPPWEDDTPVEEVKEPPVPEKKAPVLKKKDPAPAAAKARDPSPSKAAPKKVEPS  333

Qy  116  PKITTAKPI-NPRPSLPPNSDTSKETSLTTVNKETTVETKETTTTNKQTSTDGKEKTTSAK  174
         ||:| |:|           |||:   : |: |:| :| :||:| :
Db  334  SPVVPPTPVKNPVKYKPPWEDDEPAEEVKKKPSAPEKKTPVLKRKEPEPSSTTPSSDPS   393

Qy  175  ETQSTEKTSAKDLAPTSKVLAKPPPKAE------TTTKGPA                     213
         |  :|  :|     | |  | |   |            | |
Db  394  PKKAAPAVKPRDSSPKKAIPLQADPKAQEVPPTPVKNPVKKYKPPWEVDDEDPVEEVKQP   453

Qy  214  KEPTTPTTP----KEPASTTPKEP---TP-TTIKSAPTTPKEPAPTTTKSAPTTPKEPAP  264
         ||  :| |    |||| ||  ||    |  | ||:||| |||| || |::|||| |||
Db  454  EAPAKKTPVLKRKEPAAKDTAKPATSKTPETPEKKDPVKPRDSSPKKVAAKPDSAQAPA   512

Qy  265  TTTKEPA--------------PTTPKEPAPTTTKEPAPT----TTKSAPTTP         298
         |||||               | ||||| ||||  ||       || ||||
Db  513  TPVKNPVKKWRPPWEDDEDETPADDVSKPTDAKKTPSLAKKDPAPAKKESLKPKADTKAPAKP  572

Qy  299  KEP-----APTTPKRPAPTTPKE----PAPTTPKEPTPTTPKEPAPTTKEPAPTTPK-EP  348
         |||       ||  |  || ||      | | ||||| | ||||| || ||||| | ||
Db  573  RDPSPKKVAPTAPEKKTPVLAKKEPAGPADSKTKEPEKSKPRDSSPKKAVPAKPVPKTEV   632

Qy  349  APTAPKPAP-----TTTKEPAPTTPKEPAPTTTKEPAPTTP----SPTTPKEPAPTTKSAPT  398
         || |  |         ||||| | ||||| || |||||         || ||||| ||| |
Db  633  APAAVKKPEPISKPKDTAPKKAEPNSPVVP-PTPVKNPVKKWKPPWEDDDAPAKPVSLPE  691
```

```
Qy  399  TTKEPAPTTTKSAPTTP-KIPSPTTTKEPAPTTPK---EPAPTTPKKPAP--TTPKEPAPT  453
         || |   |  |  ||    |: |   | |  | ||    | |  ||  | |     |  |
Db  692  PEKK-TPVLAKKAPTKPDSEAAADPVSGPGSSKDPKLLAKKAPVKPRDPSPMKAVPIKPAPK  750

Qy  454  TPKEPAPTTTKKPAPTA-PKEPAPTTPKETAPTTPK-----EKLAPTTPKEPA          507
         || |||| |:|||||||  | | || |  | ||            ||| |||
Db  751  T--EVPPAVVKKPEPVAKSRDPESPKKAK-AEPNSP--VVPPTPVKNPVKKWKPPWEDDDA  805

Qy  508  PTTPEELAPTTPEEP------TPTTPEEPAPTTPKAAPNTPKEPAPTTPKEPAPT-PKE  561
         | ||  |  || |        | ||  | |  ||  |     | || ||  |
Db  806  PAEPVVNPVPEPEKKTPVLAKKTPVKPRDPSPKKAVPAKPSTKTDAPPVSVKPEPVSKPKE  865

Qy  562  PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTT-TTKEPTSTTSDKPAPTTP  620
         | | | |    || |  |    ||||| | | | |     |      | |::|| | || |
Db  866  PSPKKAEPNSPVVP----PTPVKNPVKW-KPPWEDDDEPTEEVKPSE--PEKKTPVLA   918

Qy  621  KGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPT------TPKKPAP-----KEL  668
         |  || ||||| |||  |  | | |    |  ||||          |  |        |
Db  919  K-KEPEKPKD-APKVAAKPRDPSPKKAVPE--KEPAKVAAKPRDLSPKKAIPIPANTQEA  974
```

RESULT   5
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR01
C;Species: Saccharomyces cerevisiae
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Accession: S48478
A;Molecule type: DNA
A;Accession: A26877
A;Residues: 1-242 <YAM>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009.
A;Molecule type: DNA
A;Accession: A26877
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A;Molecule type: DNA
A;Accession: B26877
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
A;Molecule type: DNA
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.

```
Qy  669  APTTTKGPTS------TTTSDKP-----APTTPKET----APTTPKEPAPTTPKKPAPT  711
         || | ||         ||  | |     ||  ||      || ||| |  || ||| |
Db  975  PPTPVKNPVKKWKPPWEDDDEPAEPVSAPEPEKKTPVLAKKAPAKPRDP---SPKAAPV  1031

Qy  712  TPEPPPTTTSEVSTPTTTKEPTTIHKSP----DESTPELSA-EP---TPKALENSPKEPG  763
         |  | | || |  | ||||    |         |  | | |   |       | | ||
Db 1032  AAK-PDPKIPEV-PPTPVKNPVKKWKPPWEDDDEPSEPVSAPEPEKKTPVLAKKAPTKPA  1089

Qy  764  V-------------------PTTKIPAAT-----KPEMTTTAKDKTTTERDLRTTPETT-TAAPK  802
                             | || |:|         |:|| | |::|| ||  |  |
Db 1090  TKPDSEAAADPVSGPTSKDPKLSKKAPVEKKPKPTTDKPDKDKLKPSPAKKPEKAPEPAAPK  1149

Qy  803  MTKEPAPTTTEKTTESKITATTTQVTSTTTQDTTPFKITTPLTKTTTLAPKVTTTKTTTTE  862
         |||    |||  |  :  | |  ||  ||| |    ||:  |   ::  |||  |||||
Db 1150  KWKPVWDDDDDEPEADFTVPAPSKKPDTEDPADPLG-------GPKTIKDPK-------   1193

Qy  863  IMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPTSTKKPEKTMPVRKPK--------  913
         | || |  | |   | |||  | | |  | |  |    |     | | ||
Db 1194  -LNKKAPAEKPTEK------PKPKEVSKEPPKPTEPPKP-AAPKKWKPPWEDDDAPAKPVSLPE  1243

Qy  914  --TTPTPRKMTSTMP                                              926
            | | || |  |
Db 1244  ADFTMPAPKKPDTEDP                                            1259
```

FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A;Reference number: S27281
A;Accession: S27281; MUID:89031230
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A;Reference number: JC6123; MUID:96323237
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match   11.8%; Score 798.5; DB 1; Length 1367;
Best Local Similarity   27.9%; Pred. No. 4.5e-28;
Matches 313; Conservative 108; Mismatches 523; Indels 177; Gaps 50;

Qy   44 DCDAQCKKYDK---CCPDYESFCAEVK-DNKKNRTKKKPPTPKPPVVDE--AGSGLDNGDF 97
               | | | | | | ||| || :| | | :||| |||||| ||| | :
Db   174 DLSTGCNNYDNQGHSQOTDFPGFYWNIDCDNNCGGTKSSTTSSESSTTSSTSESSSTT 233

Qy   98 KVTTPDTSTTQHNKVSTSPKITTAKPINP-------RPSLPPNSDTSKETSLTVNKETT 149
            : || | | | | | | | | ||
Db   234 TSSTSESSTTSSTSESSSTTAPAPTTTTSCTIKEKPTPPTTTSTTSESSTT 293

Qy   150 VETKETTTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETT--TKG 207
            | | | :|| | | :||| | | || | | ||
Db   294 PCTKKKTTTSK-TCT---KKTTTPVPTPS--SSTTESSSAPV----------PTPSSTTESSSA 342

Qy   208 PALTTPKE------PTPTTTPKEPASTTPKEPPTTTKSAP-----TTPKEPAPTTTKSAPT 257
           | |||| | | || || || | ||||| || | |
Db   343 PVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPV 402

Qy   258 TPKEPAPTTTKEPAPTTPKEPAPTTTPKEPAPTTTPKKPAPTTPKEP 317
           |||| || || || || || ||| || ||
Db   403 TESSSAPVTSS----TTESSSAPVTS----STTESSSAPVTSSTTESSSAPVTSSTT 451

Qy   318 APTTPKEPTP----TTPKEPAPTTPKEPAPTAPTTAPRKPAPTTPKEPAP----TIPK 370
           |||||| |||| || || || | | ||| ||||
Db   452 ESSSAPVPTPSSSTTESSSAPVT---SSTTESSSAP-VPTPSSSTTESSSAPVTSSSTTES 507

Qy   371 EPAPTTTKEPSPTTPKEPAPTTTKSAPTTPKEPAP---TTTKSAPTTPKEPSPTTPKEP 426
           | |||| |||| |||| || |||| | | |||| |||| |||
Db   508 SSAPVPTPS--SSTTESSSAPAPTPSSTTESSSAPVTSSTTESSSAPVTSSSTTESSSAPVTPSSSTTE 566

Qy   427 AP-----TTPKEPAPT-TPKKPAPTTTPKEPAPTAPKEPAPTTPKE 481
           || |||||| ||| | || |||| | | | ||||
Db   567 TPVTSSTTESSSAPVPTPSSSTTESSSAPV----PFSSSTTESSFSTGT-TVTPSSSSK 622

Qy   482 TAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPT--TPEEPPTTTPEPAPT- 535
           |||||| ||||| | ||| |||| | | | || || |
Db   623 SAPVT-----SSTTESSSAP-VPTPSSSTTESSSAPVPTPSSSTTE 677

Qy   536 KAAAPNTRKEPAPTTPKEPAPTTPKETAPT-TPKGTAPTTLKEPAPTTKKP 594
           | |||||||||| |||| || || || ||| | |||| || |||
Db   678 SSSAPVT-----SSTTESSSAPV----SSTTESSSAPVTPSSSTTESSSAPVTPSSSTTE 729

Qy   595 APKELAPTTTKEPSTISDKPAPTTPKGTAPTTPKEPAP-----TTPKEPAPT-TPKGTAPT 650
           ||| |||| |||| || | |||| | |||||||| || |||| || | | |||
Db   730 TESSSAPVPTPS-----SSTTESSSAPVTSSTTESSSAPVTPSSSTTE 788

RESULT 6
S49915
extensin-like protein - maize
C;Species: Zea mays (maize)
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
R;Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A;Description: Pex genes: pollen-specific genes with extensin-like domains.
A;Reference number: S49915
A;Accession: S49915
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1188 <RUB>
A;Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118.

Query Match   11.5%; Score 784; DB 2; Length 1188;
Best Local Similarity   29.9%; Pred. No. 1.7e-27;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

Qy   78 PIPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSL----PPN 133
           | |||||| |||||| | | ||||| | || | | || |
Db   460 PTPHSPPAD--------DYPPTPVPGKSPPATSPSPQVQPPAASTPPPSLVKLSPPQ 510

Qy   134 SDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAPTSKV 193
             ||| |||| | | | || || ||| || || | ||||
Db   511 APVG---------SPPPVKTTSPPAIG---SPSPPPVSVV 541

Qy   194 ------LAKPTPKAETTTKGPALITPKEPTPTTPKEPAPTTKSAPTTPKEPA 248
             || | ||||| || | | | |||||| |
Db   542 SPPPVKSPPPAPVGSPPPEEKSPPPAPVAVASPPPPVKSP---PPPTLVASPPPPVKSPP 599

Qy   249 PTTTKSAPTTP---KEPAPTTTKEPAPTTPKEPAPTTTKSAPTTP--KEPAPTT 305
           | || || ||||||| |||| || |||| | || | |||
Db   600 PPAPVASPPPPVKSPPPPTPVVASPPPAPVASPPPMKSPPPPPPVVSSPPPPEKSPPPPP 659

Qy   306 PKKPAPTTPKEPAPTT-----PKE---PTPTTPKEPAPTTPKEPAPTTKKPAPTTPKKP 356
           | | |||| || | | | ||| | || | ||| | | |
Db   XXX ...

```
Db   660 PAKSTPPPEEYPTPTSVKSSPPPEKSLPPPLIPSPPQEKPTPTPSKP-PSSPEKP 718
         :  ||        |||
Qy   357 APTTPKEPAPTTPKEPSPTTPKEPAPTTKSAPTTPKEPAPTTKSAPTTPK 416
Db   719 SP--PKEPVSSPPQTP----KSSPPAPVSSPPPTPVSSPPALAPVSSPP--- 768
Qy   417 EPSPTTPKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKKPAPTTPKEPAP 476
Db   769 PPAPLSSPPPAPQVKSSPPPAVQVSSP--PPAPKSSPPLAPVSSPPQVEKTSPPPAPLSSPP 827
Qy   477 TTPKETAP-----TTP---KKLTPTTPEKLAPTTPEKAP----TTPEELAPTTPEEPT 523
Db   828 LAPKSSPPHVVVSSPPPVVKSSPPPAPVSSPPLTPKPASSPPAHVSSPPPEVVKPSTP--PA 885
Qy   524 PTTPEEPAPTTPKAAAPNTPKE-PAPTTPKEPAPTTPKETAPTTPKGTAPTTPKGTAPTTPSV 582
Db   886 PTTVISP-PSEPKSSPPPTPVSLPPPIVKSSPPPAMVSSP--PMPKSSPPVVVSSPPPT 943
Qy   583 LKE---PA----PTTPKK---PAPKELAPTTKE--PTTSTSDKPAPTTPKGTAPTTP 628
Db   944 VKSSPPPAPVSSPPPATPKSSPPPAPVNLPPPEVKSSPPPTPVSSPPPA---PKSSPPPAP 1000
Qy   629 -KEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTPKGPTSTSDKPAPTTP 687
Db  1001 MSSPPPPPEVKSPPPPAPVSSPPPAPVSSPP---PPVKSPPPAPVSSPPPPV 1057
Qy   688 -TPKETAPTTPKEPAPTTPKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTPEL 746
Db  1058 KSPPPPAPISSPPPPVVKSPPPAPVS--SPPPPVKSPPPAPVSSPPPPIKSPPPPAPVS 1115
Qy   747 SAEPTPKALENSPKEPGVPTTKPAAATKPEMTTTTAKDKTTTERDLRTTPET 796
Db  1116 SPPPAP------VKPPSLP-PPAPVSSPPPPVVTPAPPKKEEQSLPPPAES 1158
Qy  1116 SPPPAP----- 1158
```

RESULT 7

T30826

nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N;Alternate names: alpha-NAC protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C;Accession: T30826
R;Yotov, W.V.; St-Arnaud, R.
Genes Dev 10, 1763-1772, 1996
A;Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A;Reference number: Z20889; MUID:96312450
A;Accession: T30826
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2187 <YOT>
A;Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1
C;Genetics:
A;Gene: Naca
A;Map position: 10
A;Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A;Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C;Keywords: alternative splicing; DNA binding; transcription factor

```
Query Match     11.4%;  Score 776.5;  DB 2;  Length 2187;
Best Local Similarity   27.6%;  Pred. No. 6.6e-27;
Matches  290;  Conservative 119;  Mismatches 416;  Indels 227;  Gaps 48;

Qy    57 PDYESFCAEVKDNKKKNRTKKRPTPKKPPVVDEAGSGLDNGDFKVTPDTSTTQOHNK----- 111
Db   809 PKVDPIMSDVTPTPSKKTISATAVPK 847
Qy   112 VSTSPKITTAKPINPRPSLPPNSDTSKETISL----TVNKET----- 163
Db   848 TSLSPPKAPVAPSNEATIVPTEIPTSLKNALAAATPKETIATSIPKVTSPSPSQKITKSVS 907
Qy   164 TDGKEKTISAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTKGPALTTPKEPTPTPKE 223
```

```
Db   908 LKGAPAMTSKKAT--EIAASKDVSPSQ--FPKEVPLLQHV--PPTSPPPKSPVSDTLSG 959
Qy   224 PASTTPKEPEPTTPTIKSAPTTPKE-PAPTTIKSAPTTPK------EPAPTTTKEPAPT 273
Db   960 ALTSPPPKGPPATLAETTPTYPKKSPKPAASKKTPATPSPEGVTAVPLEIPPCSKKAPKTA 1019
Qy   274 TPKEPAPTTPKEPAPTT--TKSAPTT------TPKEPAPTTPKK-P 309
Db  1020 APKESSATISSSKRAPKTAVSKEIPSKGVTAVPLEISLPLKETSKSATPGEKSASSPKRSP 1079
Qy   310 APTTPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPAPKKPAPTT--PKEPAP- 366
Db  1080 KTAGPKE---TEPGGVTAVPPEISLPPKETPQNATPNESLAASSQKRSPKTSVPKETPG 1136
Qy   367 --TTPKEPAPTTPKEPSPTTPK-----EPAPTTIKSAPTTPKEPAPTTI-KSAPTTPSV 416
Db  1137 GVTAMPLEIPSAPQKAPKTAVPKQIPPPEDAVTILAGSPLSPKKASKTAAPKEPATPSV 1196
Qy   417 -----EPSPTTTPKEPAPTTPKEPAPTTPKKPAPTT--PKE------PAPTTP 455
Db  1197 GVIAVSGEISPSPKKTSKTAAPKENSATLPPPKRSPKTAVPKETPAPTTPKKP 1256
Qy   456 KEPAP----TTTKKPAPTIAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEK-PA 507
Db  1257 SPPPASKGVPVTLLTPKGAPNALAE-SPASPKKVPKTAAPEETSTTP------SPQKIPK 1309
Qy   508 PTTPEELAPTTPEEPTPTT------PEEPAPTTPKAAANTPKEPAPTTPKE-PAPTT 558
Db  1310 VAGPKEASATPPSKKTPKTAVPKETSAPSEGVTAVPLEIPPSPRKAPKTAAPKETPAPS- 1368
Qy   559 PKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTP------KKPAPKE 598
Db  1369 -PEGATTAPVQIPSPRKGSKKAGSKE-TPTTSPPEGVTAAPLEIPISSKKTSKMASPKE 1426
Qy   599 LAPTTTPKEPTS-----TTSDKPAPTTPKGTAPTTPKGTAPTTPKE-PAPTTPKGTAPTT 651
Db  1427 TLIVTPSSKKLSQTVGPKETSLEGATAVPLEIPPSHKKAPKTVDPKQVPLITPSPK-DAPTT 1485
Qy   652 LKEPAPTTPKKPAPKELAPTTPKGPTSTTSDKPAPTTTPKETAPTTPKEPAPTTPKKPAPT 711
Db  1486 LAE-SPSSPKK-APKTAAPPSER-VTTVPPEKPA-TPQKASGTTPASKVPVPAETQEVAVS 1541
Qy   712 TPETP-----TPTTSEVSTPTTTTK-----EPTTIHKSPDESTPELSAEPTPKALENSPK 760
Db  1542 SRETPVTPAVPPVKNPSSHKKITSKITIELKEAPATLPPSPTPSKDTSSKKAPRT--SARK 1599
Qy   761 E-PGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTPARPKMTKETATTTEKTTESKI 819
Db  1600 EPPASPSIK-PVTT--SLAQTAPPSLQKAPSTTIPKENLAAPAV---LPVSSKSPAAPA 1652
Qy   820 TATTTQVTSTTTQDTTPFKITTLIKTTTLAPKVTTTTKITTTTEIMNKPEETAKPKDRATN 879
Db  1653 RASASLSPATAAPQTAPKEATTIPSCKKAAAATETPIETSTAPSLEGAPKETSE----TSV 1708
Qy   880 SKATTPKPQKPTRAPKKPTSTKKPKTMPRVRKPKTPIPTTSRKMTTSTMPELNPTSRIAEAML 939
Db  1709 SKVLMSSP----PKKASSSKRASTLP------AITTLPSLKEASVLS--- 1744
Qy   940 QTTTRPNQTTPNSKLLVEVENPKSEDAGGAEGETP 971
Db  1745 ----PTATSSGKDSHISPVS-DACSTGTTTP 1770
```

RESULT 8

A35175

mucin 1 precursor, repetitive splice form A [validated] - human
mucin 1 precursor, repetitive splice form A [validated] - human
N;Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episia
ncreatic mucin; polymorphic epithelial mucin (PEM)
ncreatic mucin; polymorphic epithelial tumor antigen splice form; mucin 1 precurso
N;Contains: mucin 1 precursor, epithelial mucin splice form; mucin 1 precurso
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
C;Accession: A35175; B35886; A35887; S10572; S40293; A36735; PX0066; S10218,
R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.

J. Biol. Chem. 265, 5573-5578, 1990
A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A;Reference number: A35175
R;Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
A;Accession: A35175
A;Molecule type: mRNA
A;Residues: 1-952,1033-1344 <LIG1>
A;Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124; GB:
A;Experimental source: splice form A
A;Accession: B35175
A;Note: GenBank entries HUMEPISIA1 and HUMEPISIA2 present only the amino-and carboxyl-en
A;Molecule type: mRNA
A;Residues: 1-19,29-952,1033-1344 <LIG2>
A;Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129; GB:
A;Experimental source: splice form B
A;Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino-and carboxyl-en

J. Biol. Chem. 265, 15286-15293, 1990
A;Title: Molecular cloning and expression of human tumor-associated polymorphic epithel
R;Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel
A;Accession: A35886
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-992,1033-1344 <GEN>
A;Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870
A;Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence

J. Biol. Chem. 265, 15294-15299, 1990
A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
R;Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Eur. J. Biochem. 189, 463-473, 1990
A;Reference number: A35887; MUID:90368716
A;Accession: A35887
A;Residues: 1-19,29-1109; NID:g189598; PIDN:AAA60019.1; PID:g189599
A;Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599
A;Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated seque

Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
R;Abe, M.; Siddiqui, J.; Kufe, D.
A;Reference number: S10572
A;Accession: S10572
A;Molecule type: mRNA
A;Residues: 1-19, 29-155, 'P', 177-182, 'A', 184-212,1033-1344 <LAN>
A;Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054

A;Reference number: S10571; MUID:9027413
A;Accession: S10571
A;Molecule type: mRNA
A;Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212,1033-1344 <WRE>
A;Cross-references: EMBL:X52229; NID:g37053

A;Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
R;Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A;Reference number: S40293
A;Accession: S40293
A;Molecule type: mRNA
A;Residues: 1-142, 'Q', 144-162, 'Q', 164-168 <ABE>
A;Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543

J. Biochem. 112, 609-615, 1992
A;Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
R;Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H
A;Reference number: JX0235; MUID:9312189
A;Accession: PX0066
A;Molecule type: mRNA
A;Residues: 998-1011, 'ES',1014-1017,1018-1032, 'T',1034-1037;1038-1057 <MAS>
A;Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054

FEBS Lett. 356, 130-136, 1994
A;Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
R;Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
A;Reference number: S51026; MUID:95080414
A;Accession: A36735
A;Contents: annotation.
A;Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region ar

partial repeats. The repeat shown is defined by a SmaI nuclease sites.
C;Comment: Serine and threonine residues in the tandem repeat domain are extensively
C;Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48
C;Genetics:
A;Gene: GDB:MUC1; PUM
A;Cross-references: GDB:120705; OMIM:158340
A;Map position: 1q21-1q23
A;Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C;Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorp
C;Superfamily: polymorphic epithelial mucin
F;1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
F;1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
F;1-62/Region: mucin 1 amino-terminal non-repetitive
F;138-1017/Region: 20-residue repeats (GSTAPPAHGVTSAPDTRPAP)
F;1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F;1245-1272/Domain: transmembrane #status predicted <TRM>
F;1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predic
F;1213/Binding site: phosphate (Tyr) (covalent) #status predicted

```
Query Match     Best Local Similarity   27.8%;
Matches  310;   Conservative   78;   Mismatches  472;   Indels  256;   Gaps  56;
                Score 672;   Pred. No. 1.6e-22;   DB 1;   Length 1344;

Qy    142 LTVNKETTVETKEITTTNK--QTSTDGKEKTTSARETQSIEKTSAKD-LAPTSKVLAKPT 198
              :   ||| : :  |  ||| : : :||: :     ::||:|     |
Db     15 LTIVLTATTTAPKPATVVTGSGHASSTPGGEKETSATQRSSVPSSTEKNAVSMTSSVLSSHS 74

Qy    199 P-KAETTTKGP--ALTTPKEP----------TPTTPKEPASTTP----------KE 231
              | ||| |   |  |||  ||                  |||  ||           ||
Db     75 PGSGSSTTQGSQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

Qy    232 PTPTP------IKSAPTTPKEPAPTTT-------KSAPTTPKEPAPTTKEP---A 271
              :|||             | |||     | | |             |||         |
Db    135 PAPGSTAPPAHGVTSAPDT--RPAPGSTAPPAHGVTSA 190

Qy    272 PTTPKEPAPTTTKEPAPTTKSAPTTPKEPAPTTP--------KKPAPTTPKEP---- 317
              |||  ||||   |||   :| : :| |||   |            |
Db    191 PDT--RPAPGSTAPPAHGVT-SAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 247

Qy    318 --APTTPKEPTPTTP----------KEPAPTMPKEP------APTTPKEPAPTTAPK---- 354
              || | |||        |||   |||   | ||
Db    248 TSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 307

Qy    355 KPAPTTPKEPAPTTPK---EPAPTTTPKEPSPTTPKEPAPTTT---TKSAPTTTKEPAPTT 407
              ||| |||   |||  |   | ||| |||||||   | ::
Db    308 TSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGV 367

Qy    408 TKSAPTTPKEPSPTTTKEP-----APTTPKEPAPTTPK----KPAPTTPKEPAPTTPK-- 456
              | ||  :| |     |||     | |||   ||||  |    | |||   ||||  |
Db    368 T-SAPDT--RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 424

Qy    457 ---EPAPTTTKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPE- 512
              | |   | |   | |||   ||||  |      ||   |
Db    425 HGVTSAPDTRPAPGSTAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 484

Qy    513 ---ELAPTTPEEPTPTTP-------EEPAP---TTPKA---AAPNTPKEPAPTTPK---- 552
              || |||  |       ||||   | |||   :|:|  | |::||   ||||
Db    485 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPA 544

Qy    553 ---EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP-------KKPAPKEL 599
              || | |||   ||||   |::||   |||  |          || ||
Db    545 HGVTSAPDTRPAPGSTAP----PAHGVTSAPDTRPAPGSTAPPAHGVTS 600

Qy    600 APTTIKEPISTTSDKPAP---TTPKGTAPTTPKEPAPTTPKG----TAPTTLK 653
              |||   |: | |    |      | ||  |||   |||  ||    ||||  |
Db    601 AP-PAHGVTSAPDTRPAPGSTAPPAHGVTS----APDTRPAPGSTAPPAHGVTSAPDTRP 655

Qy    654 EPAPTTP----------KKPAPKELAPTTTKGPISTTSDKPAP--TTPK----ETAPTTP 697
              : :||           ||    :||           | | ||   ||||       :||| |
```

```
RESULT  9
T18535
high molecular mass nuclear antigen - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18535
R;Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A;Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A;Reference number: Z18955; MUID:9803440
A;Accession: T18535
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1151 <SH1>
A;Cross-references: EMBL:D88440; NID:d1171138; PID:d1025045; PIDN:BAA24137.1

Query Match       9.8%;  Score 671;  DB 2;  Length 1151;
Best Local Similarity  25.9%;  Pred No. 1.5e-22;
Matches  284;  Conservative  100;  Mismatches  419;  Indels  292;  Gaps  55;

Qy     79 TPKP-------PVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTS--PKTTAKPINPRPS 129
                |||        :||    ||  :||||||     ||  |    :  ||||||||||||
Db    113 TPPPSQGPAGTPPPSQGAGAPKGDGTAQPSGTRKSGADGKPAAQDVPKATTA-ATEARP- 170

Qy    130 LPPNSDTSKETSLTVNKETTVVETKETTTINK--QTSTDGKEKTTTSAKETQSIEKTSAKDL 187
              ::        ||: ||   ::||||||| ||  ||  ||||| ||| ||||| |||
Db    171 -------ASAASPTVPKATAEATAVTAASQSAPKANATDAAAVTAA---SQSAPKATV-EV 219

Qy    188 APTSKVLAKPTPKAETTTKGPALTTPKEPTP-TTPKEPASTTPKEP---TPTTIKSAPT 242
              |||  ||||||| ||||||:||| |||  :||:||||||||||   :|||||||||
Db    220 KPAAAAVAKEAKAVIAAAAAPKATAEAKAPAVTSPTIPCSSAEAKPLTAASPTASKA--T 277

Qy    243 TPKEPAPTT---------TKSAPTTPKEPAPTTTK---------PKEPA----------- 268
              ||||||||         |||:||||||||||||         ||||
Db    278 AEAKPVPATASLMATKVTAEAKPAPSPSVPKATTDKAVATAPAKAGPDVKPAVAVCAEA 337

Qy    269 EPAPTTPKEPAPTTTKEPAPTTTTKSAPTT----------PKEPA------PTTPKKPAPTTPK 315
              |||| ||||||||||||||||||||||||          ||||      |||:||||||||
Db    338 KPAPPPPQQLPKAAAAAAPTGTELKPATAAPPHGSPRANSHTVTVTPPNVPRAAAATVP- 396

Qy    316 EPAPTTPKEPTPTT---------PKEPAPTTPKEPAPTAPKKPRAPTTTPKEPAPTTPKE 371
              ||||||||||||||         ||||||||||||||||||||||||||||||||||||
Db    397 -TAGAVPKASTGTTPAAAAPQQPVP---KAAPVTPPSPQQAVPRAATAAA----APVTPQQ 448

Qy    372 P------APTTTKEPSPTTPKEPAPTTTKSAPTTKEPAP--TTTKSAPTTPKEPSPTTKE 425
              :       || | ||| ||||| ||||| |||||||||  |||||||||||||| |||
Db    449 PVTKAATTTNATTPPQPIPKAATTTIATTAPVTPQQPIPKAGTDAAPPPAVPKAPSDGRAAT 508

Qy    426 P-----APTTPKEPAPTTPKEPAPTTPKEPAP------------TTPKEPAPT-TTTKKPAPTAP 471
              |     || | ||| ||||||||||:||||            |||||| | |||||||||
Db    509 PGVPNAATDPQPRPPTPQSVPSAAVTEPKPQPRAAPPPSNEATPAVPSPSPNLKSPLPTIP 568

Qy    472 K------EPAPTTPK----ETAPTTPKKLTPTTPEKLAPTTPKEPAPTTPEELA------ 515
              :      |||||||    |||||||||||: |||||||||||||||||||||||
Db    569 KPVPLMALTPQPVTAQMVTQLAAATKPSPIVPKASPK-ALMTPPPPGLPRALAAAAKLLG 627

Qy    516 -PTTP----------EEPTPTTP--------EEPAPTTP-------KAAAPNTPKE---- 545
              |||||          ||||||||        ||||||||       |||||||||
Db    628 LPSSPVASAMHAKVTPRPLPASPVPMAASPASLGPDAARVALATNAASPGAKPEAAGGNG 687

Qy    546 ----PAPTTPKEPAPTTPKEPAPTTPKEPAPTT-------------TPKGTAPTT----- 581
                  |||||||||||||||||||||||||||||             |||||||||
Db    688 TLMAPMGAAANTQMAPIGAAGAAQTAPMGAAAHTHVSPMGAGGATQMSPTGAANTHMSPIGA 747

Qy    582 ------------TLIKEPAPTTPKKPAPKELAPTTTKEP-------------TSTTSDKPAPTTP 620
                          |||:|||||||||||||:|||||||             ||||:|||||||
Db    748 GGATQMSPMGAAANTQMSPMGATTTQMSPMGAAANTTQPSPMGAAAATQVTATSAGNTMQVSP 807

Qy    621 KG---TAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEP---APTTPKKPAPKELAPTTTKGP 676
              ||   |||||||||||||||:|||||:|||||||||||   ||||||||||||||||||||
Db    808 MGAATPPQTPSVGAATTP-QPSPM----GAATTLMSPMGAAATTPQ---PSPMGAVTTQPP 859

Qy    677 ----TSTTSDKP-APTTPKETT----APTTPKEP---APTTPKKP-APTTPETPP----- 717
                  |||||||| ||||||||    |||||||||   ||||||||| ||||||||
Db    860 PMAAATNTTQPPPMAAASTPQSTPMGAATTTQSSPMGAATTTQSSPMGASTPQAPPTVAGSPT 919

Qy    718 PTTSEVSTPTTTTKEPTTIHKSPDESTPEL--SAEPTPKALENSPKEPGVPTTKTPAATKP 775
              |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    920 PPPIPPSPTAQTSPQPMSKSPPPDPPKAPSAAAAQTSPAAHVANASPGV-TAVSPA---P 975

Qy    776 EMTTTPAKDKTTERDLRTTPETTTAAPKMT-KETATTTEKTTESKITATTTQVTSTTTQDT 834
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    976 IGVTEASPSADCGARLSPGPTAAATDGPKASPAATADVTEAATD--VTAAATAVPA----EA 1029

Qy    835 TPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAP 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1030 AP-------TKAKRSSSSSSSSSSSSSSSSSDSDSSSSSSSESNPASPA--- 1077

Qy    895 KKPTSTKKPKTMPRVRKPKITTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLV 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1078 ---------PAVGDGQQQMTPGAAQSVPE------VTEAAVQ--------------- 1104

Qy    955 EVNPKSEDAGGAEGE 969
              |||||||||||||||
Db   1105 EAAAAAAAAAGAERE 1119


RESULT  10
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T25697
R;Fulton, B.
submitted to the EMBL Data Library, August 1996
A;Description: The sequence of C. elegans cosmid F16F9.
A;Reference number: Z20071
A;Accession: T25697
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1229 <FUL>
A;Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9; CESP:F16F9.2
A;Experimental source: strain Bristol N2; clone F16F9
C;Genetics:
```

Qy   656 APGSTAPPAHGVTSAPDTRPAPGSTAP-PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTR 714

Qy   698 KEPAPTTP------------KKPAPTTPEPTPEPPTTSEVSTPTTTTKEPTT------- 740

Qy   741 ESTPELSAEPTPKALENSPKEPGVPTTKPAA------------TKPEMTTTAKDK---TTERD 789

Qy   773 TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPD 832

Qy   790 LRTTPETT------------TAAEKMTKETATTTEKTTESKITATTTQVTSTTTQDTTTP-FKIT 840

Qy   833 TRPAPGSTAPPAHGVTSAP----DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVT 888

Qy   841 TLKTTTLAPKVTTTKKTITTT--EIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPT 898

Qy   899 STKKPKTMPRVRKPKTTPTPRKMTSTMPELNP----TSRIAEAMLQT-TTRP---NQTPN 950

Qy   947 VTSAPDTRP--APGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 1003

Qy   951 SKLIVEVNPKSEDAGGAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDEY N 1010

Qy   1004 AHGVTSAPDTRPAPGSTAPPAH----------GVTSAPDNRPALGSTA----PPVHNVTS 1049

Qy   1011 ICNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPS 1046

Qy   1050 ASGSASGSASTLVHGTSARATTTPASKSTPFSIPS 1085

Mon Apr 29 08:35:18 2002   aa3.rpr   Page 9

```
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match        9.8%;  Score 670; DB 2;  Length 1229;
Best Local Similarity  27.0%;  Pred. No. 1.8e-22;
Matches  321; Conservative  98; Mismatches 382; Indels 390; Gaps 56;

Qy      2 AWKTLPIYLLLLLSVFVIQQVSSQELSCKG-RCFESFERGRGECDCDAQCKKYDKCCPDYE  60
            ::|  |  ||  |||   | |:|  |  | |   ||:|   | | |  | ||   | : |
Db      4 AW-VVSFAFLILGNVVQSSSLLSKTINSDGSRDFKIVKHIKKNTCTCSC---KCVPD--  56

Qy     61 SFCAEVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITT 120
            |                 | :|                |:|  |    |  | | | ||
Db     57 ------------------APSNP--------------FDVSTTISSINNDN-VDIGPS--  81

Qy    121 AKPINPRPSLPPNSDTSKETSLTVNKETT-----------VETKETTTTNKQTS---TDG 166
            :|  ||   |   | |        | | ||          |  | || |   |     | |
Db     82 -----GDSNPTGSSMFQEIEATVGGQTVKSEHNIDSSVEVEKKVTTISTDASTTNAPITG 135

Qy    167 KEKTTSA-----------KETQSIEKTS----AKDLAPTSKVLAKPTPKAETTTKGPALTTP 213
            | |: |            ||:|  ||: |    | ||  |  |   |  | ||    |  | |
Db    136 KDSTTPEIITGIVVTNSKSSVTDMSTTRPSTTLSPTTTELITSP---ETLVSTDSSTST 191

Qy    214 KEPTTPTTPKEPASTTPKEPTPTTIIKSAPTTPKEPAPTTTKS------APTTPKEP 262
            || ||| | | |||||  | |||   ||||||||| || ||  |      | | || |
Db    192 EQTSPDNTTEIAS--PMETNTTT--EATTTSVEPSVSTLASEDETTVTAIAESTTTVIAE 247

Qy    263 APTTTPKEP---APTTPKE---PAPTTTKEPAPTTTKSAPTTPKEPAPTTTPKKPA--TT 313
            ||||||| |   |||||||   |   ||||| || |  ||||| || | || |  |   ||
Db    248 VSTTTEEPTTTAESTTTKKSTTKAPATTEEPTTTTEEVTTTAETSTTTSSETTSTEKPTT 307

Qy    314 P----KEPAPTTPKEPT----PTTPKEPAPTTPKEPAPTTAPKKPAPTTPKEPAP 366
            |    | | | || | |    |  | | | | ||| ||  ||  || | | |||||
Db    308 PLIDNKIAGPATSKPETTHFPVTGTTNFDTATETPVAKSEDKMTLSKTPAATETTQQTT 367

Qy    367 TTPKEPAPTTTKE-----PSPTTPKEPAPTTTKSAP----------TTTKEPAPT---- 410
            |||| | | | |      |  |||| || ||  |||          ||| | |||
Db    368 EVTDGPEKETTKNVSIEIPITTVPLVETTSTTSTASKESDGPHTTLKIKVTTADSDSTES 427

Qy    411 APTTTPKEPSPTTTKE-----PAPT------TTTPKKPAP--- 444
            |||| |  |  | ||       | |       | | |||||
Db    428 ATTVKPFNEEETTTKSHVVPKPTKKGTVKVTPKLELSFDEPTEIT-KAPHPGKLLEKKTYH 486

Qy    445 --TTPKE-----------EPAPTTPKKPAP-- 473
              ||||            | | || |||||
Db    487 FVLSDNFARYSEAKENDDYNHLDYNHYREAKEPTTTEESSTTEEVTTTEEPANTGNPPTT 546

Qy    474 PAPTT-----PKETAPTTPKKLTPTTPEKLA-----PTTPEKPAPTTPEELAPTTPEEPTPTTP 527
            |  ||     || |||||| |  |||| | |     |  | | | |||  | ||| | |
Db    547 ENPTTTEEQPTSTAESTTTALPFTTEQTVTTEEPTTAEKSTATQ--KPTTTQESVST-- 600

Qy    528 EEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPA 587
            || | | |      |    | | ||  || || ||  ||  | | || ||||
Db    601 -EKTSTTKKA--STTEE---PTTTDEPTTTT--ESSTTGKATTPELSTTSEETTTTELK 651

Qy    588 PTTPKKPAPKELAPTTTKEPTSTTSDKPAP-----TTPKGTAPTTTPKEPAPTTPKEPAPTT 643
            | || |||      | |||| ||  |          ||  |||| ||  | | | | ||||
Db    652 ITTE-----GSTTTEEPTTTAIFAEASTGIITTDEETTTSTTSTTPEITSTKE--IVT 701

Qy    644 PKGTAPTTLKEPAPTTP-----KKPAPKELAPTTTKGP--- 676
            |  |||| | | |         | | || ||  | | |
Db    702 ESAITQTSVSVVESSTPRQLPERWKAIVNKFKIHNLEVLKEKRLLKEKESTSTGSDSSE 761

Qy    677 TTTTVVAENIDEVTTEKEKVVQTTPITTTEKSTTQEETTTTTTTTEKTTSKTTEKPTTSE 722
            |  |  |         | |     |  |||||    ||||     || ||   | || |
Db    762 -TSTTSDKPAPTTPKEPAPTTPKEPAPTTTPKEPAPTTPPE--TPPPTSE 722

Qy    723 -VSPTTTPKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTPANTKPEMTTTA 781
            : || ||| |  |   |  |  |       |   | |     | || | | || |||
Db    762 PPVNTTATTHGRSLSPSSPHTVCTAWTSATSGILGTTHITTEPSTGTSHTTPAATTGTTQH 922

Qy    822 SAATTETTTSEPST-----TEST--- 854
            : | ||| ||| |      ||||
Db    923 STPALSSPHPSSRTTESPPSPGTTTPGHTTT----- 976
```

```
RESULT 11
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
R:Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter
A:Reference number: T45025
A:Accession: T45025
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:Experimental source: placenta
A:Cross-references: EMBL;Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A:Genetics:
A:Gene: MUC5B

Query Match        9.6%;  Score 651.5; DB 2;  Length 3570;
Best Local Similarity  22.7%;  Pred. No. 3.3e-21;
Matches  360; Conservative 107; Mismatches 538; Indels 579; Gaps 65;

Qy     19 IQQVSSQELSCKGRCFESFERGRGECDCDAQ-----CKKYD---KCCPDYESFCAEVKDN  69
            | |  |:|   |  ||  | | | |         ||  |   | | |  | | | |
Db     16 IDQV-GQVLTC----------SLETGLTCKNEDQTGRFNMCFNYNVRVLCCDDY-SHCFSTLAT 567

Qy     70 KKNRTKK-----------KPTTKPPVVDEAGSGLDNGDFKVT--TPDTSTTQHNKVSTSPK 117
            |||  ||           |  |  ||   | ||  | |       |  || | ||  |
Db    568 SSTATPSSTPGTTWILTKPTTIATTTPASTGSTATASSTQATAGTTPHVSTT----ATTPT 622

Qy    118 ITTTAK-------PINPRPSLPPNSDTSKETSLTVNKETTVET----KETTTTNKQTST 164
            ||| |        | | | || |  | | |||  | |||| |     | | || | ||
Db    623 VTSSKATPFSSPGTATALPALRSTATTPTTATSFTAIPSSSSLGTTWTRLSQTTTPMATMST 682

Qy    165 DGKEKTTSAKETQSIEKTSA-----------KDLAPTSKVLAKPPP---- 199
            |  |  | ||   ||||||            | ||  | | |
Db    683 ATPSSTPETVHTSTVLTTTTATTTGAATGSVATPSSTPGTAHTTKVLTTTTGFATPSSSSP 742

Qy    200 -KAET-------------TTKGPALT------TPKEPPTTTP------KEPASTTPKE 231
            :|| |             ||| | |       ||| | |||      | | ||||
Db    743 GRARTLPVWIISTTTTPTTRGSTVTPSSIPGTTHTPTVLTTTTTVATGSMATPSSSSTQTS 802

Qy    232 PTTPTTIIKSAPTTPKEPAPTTTKSAP-----TTPKEPAPTTT-KEPAPT 284
            | | |  | ||| ||| | | || |      ||| | | |||  || | |
Db    803 GTPPSLTTTTTATTTATGSTTNPSSTPGTTPIPPVLTTTTTPANISSTVTPSSALGTTHT 862

Qy    285 EPAPTTTKS-----APTTP-----APTTP--- 314
            | | || |     |||||     | |||
Db    863 PPVNTTATTHGRSLSPSSPHTVCTAWTSATSGILGTTHITTEPSTGTSHTTPAATTGTTQH 922

Qy    315 KEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTAPKKPAPTTPKEPAPTT---- 372
            |||| | | | | ||| | | || | ||  |    | | ||  || | | ||||
Db    923 STPALSSPHPSSRTTESPPSPGTTTPGHTTT-----ATSRTTATATAPPSKTRTSTLLPSQP 976
```

```
Qy    782 KDKTTFERDLRTTPETTTA----------------------------------APK 802
            | |   | |  |  | |||                                  |
Db    855 ETTTTTSAE---TSETTTSESAAFITGESPENTALQSSSQKSEENESSAAEKPGAARRDFVPK 911

Qy    803 MTKETATTTEKTTESKITATTTQ-VTST---TTQDTTPFKITTLKTTTLAPKVT---TTK 855
            | | ||| | |   | | | |||  | |    | | | | | | | | ||    |  |||
Db    912 KHKTTVKPAETTSAVAASTTTTEPITTTEKSTTLETTPTIEATTLNEVTTGPAFVTGAPVDE 971

Qy    856 KTITTTEIMNKPEETAKPKDRATNSKAITTKPEQ-----------------------KPTKAPKKPT 898
            |   ||      |   |          | | |||                       |   | | |
Db    972 TTINTLELLSK----------INNTQISQPKPTIDISKTDALSSLIGSFTKAPMAPT 1021

Qy    899 STKKPKTMPVRVRKPKTTPTPRKMTSTMPELN----------------------PTSRIAEA 937
            |  | | |     ||  |  | ||  | |                        |   | |
Db   1022 I-----------------HTTTDAAFVTATEASLNDGSDKKIIDEAQPTDEIRRA 1059
```

```

```
Qy   373 --APTTTK-----EPS--- 381
             || || ||
Db   977 TSAPITTVVTMGCEPQCAWSEWLDYSYPMPGPSGGDFDTYSNIRAAGGAVCEQPLGLECR 1036

Qy   382 ---PTTPKEP 388
                                                                |:||
Db   1037 AQAQPGVPLRELGQVVECSLDFGLVCRNREQVGKFKMCFNYEIRVFCCNYGHCPSTPATS 1096

Qy   389 APTTTKSAPTTT--KEPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKKPAPTT 446
           : | | | || || : :| :||:| || | : || | |:
Db   1097 STATPSSTPGTTWILTELTTTATTTESTGSTATPTSTLRTAP--PPKVLTTTATTPTVTS 1154

Qy   447 PKEPAPTTPKEPAPTTTKKPA-------PTAPK-EPAP--------------TTPKETAP 484
            | || | | || ||| | | ||| |
Db   1155 SK----ATPSSSPGTATALPALRSTATTPTATSVTPIPSSSLGTTWTRLSQTTTPTATMS 1210

Qy   485 TTPKKLTPTTPE------KLAPTTPEKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAA 538
           | || | | || : || | ||| || ::
Db   1211 TATPSSTPETAHTSTVLTATATTTGATGSVATPSSTPGTAHTTKVPTTTTTGFTATP-SS 1269

Qy   539 APNTPKEP-----APTTPKEPAPT-TPKEPAPTTPKETAPTTPKGTAPT-TLKEPAPTTP 591
           :| | | ||| | || || | || | | :: |: :|
Db   1270 SPGTALTPPVWISTTTTPTTRGSTVTPSSIPGTTHTATVLTTTTTTVATGSMATPSSSTQ 1329

Qy   592 KKPAPKELAPT--------TTKEPTSTTSDKPAP-------TTPKGTAPT-TPKEPAPTT 635
               | | | :| |:|| :| | ||| |: | || ||
Db   1330 TSGTPPSLTTTATTITATGSTTNPSSTPGTRPIPPVLTTTATTPAATSSTVTPSSALGTT 1389

Qy   636 PKEPAPT--TPKGTAPTT 651
              | | || | ||
Db   1390 HTPPVPNTTATTHGRSLSPSSPHTVRTAWTSATSGTLGTTHITEPSTGTSHTPAATTGTT 1449

Qy   652 ------LKEPAP---TTPKKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTT--PKEP 700
                 | | | || |:| ||| | |: || | || :| :| | |
Db   1450 QHSTPALSSPHPSSRTTESPPSP----GTTTPGHTTATSRTTATATPSKTRTSTLLPSSP 1505

Qy   701 -- 700

Db   1506 TSAPITTVVTMGCEPQCAWSEWLDYSYPMPGPSGGDFDTYSNIRAAGGAVCEQPLGLECR 1565

Qy   701 --APTTPKKP 708
                                                                |:||
Db   1566 AQAQPGVPLRELGQVVECSLDFGLVCRNREQVGKFKMCFNYEIRVFCCNYGHCPSTPATS 1625

Qy   709 APTTPETPPPTT---SEVSTPTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVP 765
           : || : | || :| :| || | | ||| | || | : |
Db   1626 STATPSSTPGTTWILTEQTTAATTTATTGSTAIP-SSTP--GTAPPPKVLTSQATTPTAT 1682

Qy   766 TTKTPAATKPEMTTT------AKDKTTERDLRTTPETT-------------------TAA 800
           ::| ::: | || |:| | :| |
Db   1683 SSKATSSSSPRTATTLPVLTSTATKSTATSFTPIPSSTLGTTGTSQNRPPHPMATMSTIH 1742

Qy   801 PKMTKETA-TTTEKTTESKITATTTQVTST-TTQDTTP------FKITTLKTTTLAPKVT 852
           | | || |:| ||: ||||: ||: :| :|| ::|| ||| |
Db   1743 PSSTPETTHTSTVLTTK----ATTTRATSSMSTPSSTPGTTWILTELTTAATTTAALPHG 1798

Qy   853 TTKKTITTTEIMNKPEETA-----------KPKDRAT--NSKATTPKPQKPTKAPKKPTS 899
           | | || |: :| || ||| | | || : |
Db   1799 TPSSTPGTTWILTEPSTTATVTVPTGSTATASSTRATAGTLKVLTSTATTPTVISSRATP 1858

Qy   900 TKKPKT---MPRVRKPKTTPTPRKMTSTMPELNPTSRIAEA--MLQTTTRPNQTPNSKLV 954
           : | | :| :| |||| :|: |:| : | | || | | ::
Db   1859 SSSPGTATALPALRSTATTPTATSVTAI-----PSSSLGTAWTRLSQTTTPTATMSTATP 1913

Qy   955 EVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVP---NQGIIINPMLSDETNI 1011
              |:: | | | || : :|| | | | | :
Db   1914 SSTPETVHTSTVLTTTATTTRTGSVATPSSTPGTAHTTKVPTTTTTGFTATPSSSPGTAL 1973

Qy   1012 CNGKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPPSPARRI-------------------- 1051
               || || | || ::| | |
Db   1974 T--PPVWISTT----TTPTTRGS---TVTPSSIPGTTHTATVLTTTTTTVATGSMATPSS 2024

Qy   1052 -TEVWGIPSPIDTVFTRCNCEGKT 1074
            |: | | : | | | |
Db   2025 STQTSGTPPSLTTTATTITATGST 2048

RESULT 12
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34513
R;Favello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of C. elegans cosmid ZK783.
A;Reference number: Z21536
A;Accession: T34513
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3507 <FAV>
A;Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A;Experimental source: strain Bristol N2; clone ZK783
C;Genetics:
A;Gene: CESP:ZK783.1
A;Map position: 3
A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/
 3504/1

  Query Match   9.4%; Score 641.5; DB 2; Length 3507;
  Best Local Similarity   23.4%; Pred. No. 8.8e-21;
  Matches 265; Conservative 167; Mismatches 444; Indels 257; Gaps   49;

Qy   25 QELSCKGRCFESFERGRECDC-----DAQ--------------CKKYDKC-----CPDYE 60
           | | | | : : : :| |: | |:| | |
Db   1853 QSLQISGLCAQKNDCNKHAECIDIHPDSHFCSCPDGFIGDGMICDDVDECNNAGMCDDEN 1912

Qy   61 SFC---------------AEVKDNKKNRTKKKPTPKPPVVDEAGSGLDN-GDFKVTT--- 101
           : | :| : ||:| :   :|| | | | | ||
Db   1913 TKCENTIGSFNCVCLEGFKKVDEKCVVDEKKQPNREKIEIDEENSSSSNSGQEKPTTKGI 1972

Qy   102 -PDTSTTQHNKVSTSPKITTA-------KPINPRPSLPPN---SDTSKETSLTVNKETTV 150
              || | : | :||: | : | | | | | | | : :| ||
Db   1973 VSSTSATSSESTTAEPHVTTSISSTTSTKDMTSSKS-PENVTMSSESPEVSTSSSKSTTA 2031

Qy   151 -ETKETTTTNKQTSTDGKEKTTSAKETQSIEKTSAKDLAP----TSKVLAKPTPKAETTT 205
            || ::| :: :|::   :: | |: | ::| | | :|:: | : |: |
Db   2032 SETTVSSTPSESSSSEAPLTSSPATTTEVITESSVKSTTPKEESSSEITVKLSSKSPEVT 2091

Qy   206 KGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKS-APTTPKEPAPTTTKSAPTTPKEPAP 264
           :   ::| |: || : || |: | : | || | | | |: | | |
Db   2092 ESSVKSSPSTPS-TTSQSVTSTVPETSKSTVLSSEAPVTSTSPTEVHT-SSETKPSLSAS 2149

Qy   265 TTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPK-EPAPTTPK 323
           :|| : ||| : : | : || | : : :| |:| :   ::
Db   2150 STTGDTNSTTPSTSSLASVKSTSAPEGTSASVAPVKLSSLSPDVSQPSTKTFDATESSTV 2209

Qy   324 EPTPTTPKEPAPTTKEP-APTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSP 382
           : : |:   :| || : | : | | | : || |:   :|::|:
Db   2210 QASETSSGTSVKSTSEPESHVTKLSITSSNPSSSVPVTSPKSTPTVPE-----STEQPTS 2264

Qy   383 TTP--KEPAPTTTKSAPTTTKEPAPTTTKSAP------TTPKEPSPTTTKEPAPTTPKEP 434
           ||| : | : | || || :: :| ||| | ::| | |||
Db   2265 TTPSGQSLTPMNSNSEVLTTSEPHVLSSSLSPDVSQSSTTPNNLSESSTVE----TPKTS 2320

Qy   435 APTTPKKPAPTTPKEPAPTTPKEPAPTT---------------------TKKPAPTAPK- 472
           : : |:| : | :| : || ::|| :||:
Db   2321 SEVSLNSEEPSTTEAPTTLSPDILSTTTNNLSQSSTVSTEDRSEISSENSEKPT-SAPEL 2379

Qy   473 ---------EPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPT 523
                      :| | |: : | || :: | :   : ::   :|| | ||||
Db   2380 VTSSVTHVASSSPDVPTES--SEPDDLTGSSTENIPEASSKQTISSTPTPDTTTASEEPT 2437

Qy   524 PTTPEEP-----------APTTPKAAAPNTPKEPAPTT------------PKEPAPT--- 557
            :| | : |||:::: | | :: | | |
```

RESULT 13
I38346
        elastic titin - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: I38346
R;Label1, S.; Kolmerer, B.
Science 270, 293-296, 1995
A;Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A;Reference number: 138346
A;Accession: A57430; MUID:96026330
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-7962 <RES>
A;Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:127867; OMIM:188840
A;Map position: 2q31-2q31

Query Match    9.3%;     Score 635.5; DB 2; Length 7962;
Best Local Similarity   22.8%;    Pred No. 3.7e-20;
Matches 306; Conservative 112; Mismatches 475; Indels 451; Gaps 63;
```

```
Db  2438  KSTSMSPDLSTTSNVLSESSTTPESSS----------KSPVSSSTEGISVVSTEFSKVPESTISS  2493
Qy   558  ------------TPKEPAP------------------TTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKE   606

Db  2494  VLEEDLTKTTSPILEETTTASETSEPLIEDSLTVSVRIHELTTSSENVPKESESTTTSS  2553
Qy   607  PTSTTSDKPA----------------------TP--KGTAP---TTPK-------   629

Db  2554  ESSKPSQOEPAGILTSTVVVPISSVSVSLITASEIEATTSNPFKQCRTPIITTSPKSLVKSTT  2613
Qy   630  EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKGPT-STTSDKPAPTT   688

Db  2614  SPSTVTSSEPSESSTKRTTVSTTVSTTTPTEETTTSESLLITAAPSKPTESTTESSEAPTT  2673
Qy   689  PKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTTIHKSPDESTP--EL   746

Db  2674  PAKTSETKPSNVSSTSRKSTENVETSTSQGSGSLESSTMSSTS-----SEPETNAPAVTV  2727
Qy   747  SAEPTPKALE--NSPKEPGVPTTKTPAATK----PEMTTTAKDKTTERDLRTTPETTTA   799

Db  2728  SSEASSTTLEENSSTSSP----TSSEASVKLSSLFPESITSEAVTVSSR----APAETTM  2779
Qy   800  APKMTKETATTTEKTTESKITATTT-------QVTSTTTQDTTP-FKITT--   841

Db  2780  SSESHREISTVSSSEPSEPIPLSTTVSPNVVTASSIPSEEEPILLSSVSTSSSTPRVRLTTGT  2839
Qy   842  ---LKTTTLAPKVTTTTKKTITTTTEIMNKPEETAKPKDRATNSKATTPKPQKPTKAPKKPT   898

Db  2840  PDDLIVSVTVPSHGNRRQNITASSV---PSNSTSPIILPSES-LTTPQPPPTTTTAKPA  2895
Qy   899  STKKPKTMPRVRKPK-------TTPPTPRKMTSTMPELN------PTSRIAEAMLQTTT   943

Db  2896  TTSGKRGPPSIQPPAEMFTTPAP--------PPPSNGGYGEEETNQEEEQVTSTTT  2942
```

```
Db  5956  LPERKPAPEEVAVPVPIPKKVPEPPAPKVPEVPKRP--VPEEKKPVPVPKPEVPVPK  6014
Qy   231  EPTPTTIKSAPTTPKEPAP-------TTTKSAPTTPKEPAPTTPKE--VPEEKKPVPVPKP   277

Db  6015  KPVPEEKIPVPVAKKKEAPPAKVPEVQKGVVTEEKITIVTQREESP---PPAVPEIPKK  6070
Qy   278  PAPTTTKEPAPTTPKSAPTTPKEPAPTTPKKPAPTTP-KEPAPTTPKEPTPTT-   329

Db  6071  KVPEERK-PVPRKEEEVPPPRKVPA--LPRKPVPEEKVAVPVPVAKKAPPPRAEVSKKTV  6127
Qy   330  --------------------------------------------RKEPAPT   336

Db  6128  VEEKRFVAEEKLSFAVPQORVEVTRHEVSAEEEWSYSEEEGVSISVYREEEREEEEAEV  6187
Qy   337  TK-:------------------------------EPAPTT------PKEPA   351

Db  6188  TEYEVMEEPEEYVVEEKLHIISKRVEAEPAEVTERQQEKKIIVLKPKTIPAKIEEPPAPKVPE  6247
Qy   352  APKK-------PAPTTPKE--PAPTTPKE-APTTTKEPSPTTPKE----PAPTTTKSAP   397

Db  6248  APKKIVPEKKVPAPVPKKEKVPPKKVPEEEKKVPEBKKVPPKVIKMEEEPLPAKVTEKHMQ  6307
Qy   398  TTTTKEP--APTTTKSAP---TTPKEPSPTTTKE-----------   430

Db  6308  ITQEEKVLAVTKKEAPKARVPEEPKRAVPEEKVLKLKPKREEEPPAKVTEFRKRVVKE  6367
Qy   431  ------PK-EPAP---------------------------PAPTT----   436

Db  6368  EKVSIEAPKREPQPIKEVIIMEEKERAYTLEEEAVSVQREEEYEEEYDYKEFEEYEPT  6427
Qy   437  ----------TTPKKPAPT---TPKEPAPTTTPKEPAPTTPKKREPVPVPVALPQ   470

Db  6428  EEYDQYEEYEEREVEREYEEHEEVTTEPEKPIPVKPVPEEPVPTKPKPAPPAKVLKKAVP-E  6486
Qy   471  PKEPAPTTPKETAPTTPKKLTPTTPEK-----LAPTTPEKPAPTTPEELAPTTPEPT   523

Db  6487  EKVPVP-IPKKLKPPPPK--VPEEPKVFEEEKLHISITKREKEQVTEP--AAKVPMKPK  6540
Qy   524  PTTPEEPAPTTPKAAAP---NTPKEPAPTTPKEPA---------------   555

Db  6541  RVVAEEKVPAEEKVPVPRKEVAPPVRVPEVPKELEPEEVAFEEEEVTHVEEVLVEEEEYIHEEEE  6600
Qy   556  PTTPKEPAPTTPKETAPTTPKKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTT----   611

Db  6601  FTIEEEVVPVIPVK-VPEVPRKPVPEE-KKPVPVPRKKKEAPPAKVPEVKPEEKVPVLI  6658
Qy   612  --SDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPA--PTTPKKPAPKE   667

Db  6659  PKKEKP----PPAKVPEVPKKPVP--EEKVPVPVPKKVEAP----PAKVPEVPKKPVPEK  6708
Qy   668  L----APTTTKGPTSTTSDKPAPTTPKEPAPT-TPKE---PAPTTPKKPAPTTPETPPPT   719

Db  6709  KVPVPVPAPKKVEAPPAKVPEVPKKLIPEEKKPTPVPRKKVEAPPPKKPVPKKREPVPVPVALPQ  6768
Qy   720  TSEVSTPTTTIKEPTTIHKSPDESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTT   779

Db  6769  EEEVLPEEEIVPEEEVLPE-EEEVLPEEEEVLPEEEVLPEEEIPPEEEEVVPEEEEVVP  6827
Qy   780  TAKDKTTERDL--RTTPETTTAAPKMTKETATTTEKTTESKITATTTQVTSTTTQDTTPF   837

Db  6828  EEEEVPEEEYVLPEEVKPKVPVPAP----VPEIKKKVTEKKVVIPKKEEAPPAKVPEVPK  6882
Qy   838  KITTLKTTTLAPKVTTTTKKTITTTTEIMNKPEETAKPKDRATNSKATTPKKEEAPPAKVPEVPK   897

Db  6883  KVE--EKRIILPK----EEEVLPVEVTEEPEE------------EPISEEEIPEEP  6920
Qy   898  TSTKK-----PKTMPRVRK---PKT-TPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT   948

Db  6921  PSIEEVEEEVAPPRVPEVIKKAVPEEAPTEVPVPKKVEA--PPAKVSKKIPEEKVPVPVQKKEA  6978
Qy   949  PNSKLVEVNPKSEDAGGEAEGEETEHMLLRPHVFMPEVTPDMDYLPRVPNQGIIINPMLSDE  1008
```

```
Db    6979  PPAKVPEV----------------------PKKVPEKKVL--VPKKEAV---------- 7003
                                          |||||
Qy    1009  TNICNGKPVDGLGTTLRNGTLVAFR 1032
                                          |||||
Db    7004  --------PPAKGRIVLEEKVSVAFR 7021
```

```
RESULT 14

T11622    extensin class 1 precursor - cowpea
C;Species: Vigna unguiculata (cowpea)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C;Accession: T11622; S54155
A;Accession: T11622
R;Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
Mol. Plant Microbe Interact. 10, 95-101, 1997
A;Title: Rhizobia modulate root-hair-specific expression of extensin genes.
A;Reference number: Z17301; MUID:97155574
A;Accession: T11622
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-489 <ARS>
A;Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937
A;Experimental source: sub_species Red caloona
R;Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
submitted to the EMBL Data Library, April 1995
A;Description: A class of root-hair specific extensins involved in rhizobium/legume inte
A;Reference number: S54155
A;Accession: S54155
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 326-489 <AR2>
A;Cross-references: EMBL:X86030; NID:g791149; PID:g791150
C;Genetics:
C;Gene: Ext26G
C;Superfamily: hydroxyproline-rich glycoprotein
C;Keywords: glycoprotein; hydroxyproline
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-489/Product: extension class 1 #status predicted <MAT>
```

```
Query Match          9.3%;    Score 633;  DB 2;  Length 489;
Best Local Similarity      31.2%;    Pred. No. 2.9e-21;
Matches 149;  Conservative     33;  Mismatches 252;   Indels 44;  Gaps  7;
```

```
Qy  244  PKEPAPTTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTTKEPAPTTPKEPAP 303
         ||:| |   | |||  ||  |||| |||| | ||  |||| |||:|
Db   39  PKQTPPYVYNAPPYYYKSPPP-----PSPSPP---------------------- 83

Qy  304  TTPKKPAPTTPKEPAPTTPKEPTTPKEPAPTTPKEP---KEPAPTAPKKPAPTT 360
         :| ::| | || | | |||  || |  |:| |||    || | |||  |
Db   84  PSPSPPPPPYVYKSPPPPSPSPPPPPYVYKSPPPS--PSPPPPPYVYKSPPPPY 141

Qy  361  PKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPSP 420
         || | |||  ||| |   | |    || | |||  ||| |
Db  142  YKSPPPSPSPPPPPYVYKSPPPPSPSPPPPPYVYKSPPP--------------- 187

Qy  421  TTTKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKKPAPTTTPKEPAP 480
         | ||  ||| | | |   || |   |  || |  ||| |
Db  188  YYYKSPPPPSPSPPPPPYVYKSPPPPSPSPPPPPYVYKSPPP------- 247

Qy  481  ETAPTTTPKKLIPTTPKEKLAPTTPKKPAPTTTPEELAPTTPEEPAPTTPKAAAP 540
         | || | | | |   ||  ||| |   | ||  |:   |||
Db  248  PPPPYVYKSPPPPSPSPPPPPYVYKSPPPPSPSPPPPPYVYKSPPP 307

Qy  541  NTPKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELA 600
         :|   | |   ||  ||  ||| | | | | |||
Db  308  PSPSPPPPPYVYKSPPPPSPSPPPPPYVYKSPPPPSPPPPPYVYKSPP- 366

Qy  601  PTTTKEPTSTTSDKPAPTTPKGTAPTTPKETAPTTPKKPAPTTPKEPAPTTP 660
         | || | |   | |   ||  || | |   ||  ||| |
Db  367  ----KSPPPPSPSPPPPPYVYKSPPPPSPSPPPPPYVYKSPPPPSPSPPPP 422

Qy  661  KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKKPAPTTPETPP 718
```

```
Db  423  SPP--------PPYVYKSPPPPSPSPPPPPYVYKSPPPPSPSPPPPPYYYKSPPPPSPSPPPP 475
```

```
RESULT 15

C84672    hypothetical protein At2g27380 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84672
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: C84672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-761 <STO>
A;Cross-references: GB:AE002093; NID:g5306260; PIDN:AAD41992.1; GSPDB:GN00139
C;Genetics:
C;Gene: At2g27380
A;Map position: 2
```

```
Query Match          9.3%;    Score 632;  DB 2;  Length 761;
Best Local Similarity      31.1%;    Pred. No. 5e-21;
Matches 217;  Conservative     44;  Mismatches 354;   Indels 82;  Gaps 33;
```

```
Qy  197  PTPKAETTTKGPALTTP---KEPPT--TPKEPASTTPKEPTTPTTIKSAPTTPKEPAPTTT 252
         ||| |   | |  || |   || ||  |||| | |||  ||:||| | ||  ||| | | |
Db   69  PPPIQKPPTTYSPPIYPPPIQKPPTTYSPPIYPPPIQKPPTTYSPPIIYPPPIQKPPTTPT 128

Qy  253  KSAPTTTPKEPAPTTTPKEPAP---TTPKEPAPTTTTKEPAPTTTTKSAPTTP----KEPAPT-- 304
         |  | | || || |  ||  |    ||| | | | | |||  ||  | |  |    || | |
Db  129  YSPPIYP----PPIQKPPTSYSPPVKPPPV---QMPPPTTYSPPVKPPPV--TPKEPAPTTPKK 181

Qy  305  TPKKP---APTTPKEPAPTTP----KEPTP--TTPKEPAPTTIKEPAPT-TPKEPAPTAPK 354
         ||  |   || | |  || ||    ||  |  |||| | ||   || | |||| | | |
Db  182  PPIKPPVHKPPTPIYSPPIKPPPVHKPPTPIYSPPIKPPPVHKPPTTPIYSPPVKPPPVHK 241

Qy  361  PTPIYSPPVKPPPVHKPPTPIYSPPIKPPPTIQKPPTTYSPPI----TPKEPAPTAPK 420
         | | |  |||  |||| |||  || ||  |  || | | || |    | | | | | |
Db  242  PPTPIYSPPIKPPPVHKPPTPIYSPPIYPPPIQKPPTTPIYSPPVKPPPVHKPPTTPYSPP 301

Qy  466  PAP---TAPKEPAPT---TPKETAPTTP----KKLIPTTPEKLAPTTPEKPAPTT---PE 512
         | |   | |  | |    |||| | |      ||| | | | | |   | | |   ||
Db  302  VKSPPVQKPPTPIYSPPIKPPPVQKPPTTYSPPIKPPPVKPPTPIYSPPIKP--PPVHKP 360

Qy  421  KLPPVKPPTPIYSPPVKPPPVHKPPTPIYSPPIKPPPVKPPTP 480
         ||| | | |  ||  | ||   |  | | |  || ||  | ||
Db  421  TTPKKPAPKELAPTTTKEPTSTTTSDKPAPTTPKGTAP 625

Qy  513  ELAPTTPEEPTPTTPEEPAPT-----TPKAAAP--NTPKEPAPTTPKEPAPTTPKEPAPTTP 567
         | || | |  | |  |||| |     ||   |   |||| | ||  ||| |  ||  | |||
Db  361  PTPIYSPPVKPPPVHKPPTPIYSPPVKPPPIQKPPTPIYSPPIKPPPVKPPPLQKPPTPIYSPPI 420

Qy  355  KPAPT-TPKEPAPTTPKEPAP---TTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTIKS 410
         ||| | |||| | |  ||  |   || ||  | |||  | ||  | |  ||||| | | ||
Db  355  PPTPIYSPPIKPPPVHKPPTPIYSPPVKPPPVQTPPTPIYSPPVKPPPVHKPPTPIYSPP 410

Qy  182  TPKKP-----PPIKPPPVHKPPTPIYSPPIKPPPVHKPPTPIYSPPVKPPPVHK 241
         ||  |     ||||| |||| |||  || ||  |||| |||  || ||  |||
Db  187  PPSPSPPP

Qy  568  KETAPTTPKGTAPTTLKEPAPT---TPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAP 625

Qy  481  TYSPPVQP-----PPVQKPPTPIYSPPVKPPPIQKPPPI----PTYSPPIKPPPVKPPTPIY 532
         ||||| |      || |||| |  ||||  |  || | |    || |  || |||| |||
Db  648  ---TPIYSPPIKPPPVQKPPTPIYSPPVKPPPVQLPP-TPTYSPPVKPPPVQVPPTPIYS 703

Qy  626  TTPKEPAPTTPKEPAPT--TPKGTAPTTLKEPAPT---TPKKPAPKELAPTTPIKGPSTTSD 682

Qy  533  SPPIKP-PPVHKPPTPTYSPPIKPPPIHKPPTPTYSPPIKPPP 591

Qy  683  KPAPTTPKETAPTTPKEPAPTTPKKPAPT-TPETPPPTTSEVSTPTTTK--EPTTIHKSP 739

Qy  592  VHKPPTPTYSPPIKP----PPVHKPPTPTYSPPIKPPPVHKPP 647

Qy  740  DESTPELSAEPTPKALENSPKEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPETTTA 799

Qy  648  ---TPIYSPPIKPPPVQKPPTPTYSPPVKPPPVQLPP-TPTYSPPVKPPPVQVPPTPIYS 703
```